

## 거주지 분화에 대한 공간통계학적 접근 (II): 국지적 공간 분리성 측도를 이용한 탐색적 공간데이터 분석\*

이상일\*\*

### A Spatial Statistical Approach to Residential Differentiation (II): Exploratory Spatial Data Analysis Using a Local Spatial Separation Measure\*

Sang-II Lee\*\*

**요약:** 이 논문의 주된 목적은 국지적 공간 분리성 측도를 이용한 탐색적 공간데이터 분석을 통해 거주지 분화 연구에서 공간통계학적 접근이 가지는 의의를 논증하는 것이다. 탐색적 공간데이터 분석은 공간 데이터를 다양한 과학적·지도학적 시각화 방식을 통해 탐색함으로써 패턴을 발견해 내고, 의미 있는 가설을 수립하며, 더 나아가 공간 데이터에 대한 통계학적 모델을 평가하는 것을 주목적으로 한다. 이 연구는 국지 통계량에 기반한 탐색적 공간데이터 분석이 구체적인 연구 수행에서 실질적인 도움을 줄 수 있다는 믿음에 기반을 두고 진행된 것이다. 중요한 결과는 다음과 같다. 첫째, 이미 개발된 전역적 공간 분리성 측도로부터 국지적 공간 분리성 측도를 도출하였다. 둘째, 두 가지 유의성 검정을 위한 가정, 즉 총체적 랜덤화 가정과 조건적 랜덤화 가정에 기반한 가설검정 방법을 제시하였다. 셋째, 측도와 유의성 검정을 바탕으로 한 탐색적 공간데이터 분석 기법으로 '공간 분리성 산포도 지도'와 '공간 분리성 이레지 지도'를 제시하였다. 부가적으로 각 인구 집단 별 집중도에 대한 표준화 지표도 제시되었다. 넷째, 개발된 기법을 우리나라 7대 도시의 고학력 집단과 저학력 집단간 거주지 분화에 적용한 결과, 특히, 이변량 공간적 클러스터와 공간적 특이점을 확인하는 데 유용성이 있는 것으로 드러났다.

**주요어:** 탐색적 공간데이터 분석, 공간 분리성 측도, 국지 통계량, 거주지 분리, 공간적 의존성

**Abstract:** The main purpose of the research is to illustrate the value of the spatial statistical approach to residential differentiation by providing a framework for exploratory spatial data analysis (ESDA) using a local spatial separation measure. ESDA aims, by utilizing a variety of statistical and cartographic visualization techniques, at seeking to detect patterns, to formulate hypotheses, and to assess statistical models for spatial data. The research is driven by a realization that ESDA based on local statistics has a great potential for substantive research. The main results are as follows. First, a local spatial separation measure is correspondingly derived from its global counterpart. Second, a set of significance testing methods based on both total and conditional randomization assumptions is provided for the local measure. Third, two mapping techniques, a 'spatial separation scatterplot map' and a 'spatial separation anomaly map', are devised for ESDA utilizing the local measure and the related significance tests. Fourth, a case study of residential differentiation between the highly educated and the least educated in major Korean metropolitan cities shows that the proposed ESDA techniques are beneficial in identifying bivariate spatial clusters and spatial outliers.

**Key Words:** exploratory spatial data analysis (ESDA), spatial separation measure, local statistics, residential segregation, spatial dependence

\* 이 논문은 2005년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2005-003-B00389).

\*\* 서울대학교 지리교육과 조교수(Assistant Professor, Department of Geography Education, Seoul National University), si\_lee@snu.ac.kr

## 1. 문제 제기 및 연구 목적

도시 내에서 상이한 인구 집단들의 거주지가 얼마나 ‘공간적’으로 분리되어 있는가를 측정하는 문제는 인구 집단들이 공간을 가로질러 얼마나 상이하게 배열되어 있는가를 측정하는 것인데, 이 ‘배열’은 본질적으로 두 가지 의미를 갖는다. 하나는 두 인구 집단의 거주지 중첩도를 의미하고, 또 다른 하나는 그러한 중첩도의 분포가 보여주는 공간적 집중도를 의미한다. 진정으로 공간적인 거주지 분리에 대한 측도는 바로 이러한 ‘배열’에 대한 두 가지 개념을 모두 포함하는 것이어야 한다. 이러한 의미에서 이상일(2007)은 최근 공간통계학에서 개발된 ‘공간 연관성 측도’와 ‘공간적 카이-스퀘어 통계량’을 결합함으로써 새로운 ‘공간 분리성 측도(spatial separation measure)’를 제시하였다. 본 연구의 주된 초점은 이전 연구에서 개발된 전역적(global) 측도로부터 국지적(local) 측도를 이끌어 내고, 그것을 바탕으로 한 다양한 탐색적 공간데이터 분석(exploratory spatial data analysis; 이하 ESDA) 기법을 제시함으로써 거주지 분화 연구에서 공간통계학적 접근이 가지는 의의를 예증하는 것이다.

ESDA는 통계학이나 데이터 분석론에서 새로운 패러다임으로 등장한 탐색적 데이터 분석(exploratory data analysis; 이하 EDA)의 공간 분석적 확장이다. 최초의 주창자로 알려진 Tukey(1977)는 EDA를 ‘탐지적 연구(detective work)’라고 간명하게 정의한 바 있고, Good(1983)는 “기술적 통계학과 추론적 통계학 사이에 있는 중간적 통계학(intermediate statistics)”이라 규정하였다. 좀 더 구체적으로 EDA의 의미를 살펴보면, “다양한 시각적 재현 수단을 사용함으로써 데이터에 대한 연구자의 직관을 향상시키는 다양한 기법들의 총합”이라고 볼 수 있다(Lee, 2005, 276). 이러한 EDA의 ESDA로의 확장은 공간데이터 분석에 가담해온 학자들에게는 매우 자연스러운 것으로 인식되었다. 왜냐하면 공간데이터 분석은 소위 ‘지도학적 분석’이라고 알려져 있는 고유한 패턴 인식 테크닉을 사용해 오고 있었기 때문이다. 따라서 ESDA는 공간 데이터를 다양한 과학적·지도학적 시각화 방식을 통해 탐색함으로써

패턴을 발견해 내고, 의미 있는 가설을 수립하며, 더 나아가 공간 데이터에 대한 통계학적 모델을 평가하는 것을 목적으로 한다(Lee, 2005, 275).

이러한 ESDA의 발전과 맞물려 있는 두 가지 요소가 있는 데, 그것은 각각 지리정보시스템(Geographic Information Systems; 이하 GIS)과 국지적 통계량(local statistics)이다. 많은 학자들은 GIS가 가지고 있는 공간 데이터 조작(spatial data manipulation) 기능과 시각화 기능이 ESDA의 공리를 실현시키는데 적절하고, 거꾸로 ESDA의 성장은 ‘일반 연구 환경(generic research platform)’으로서의 GIS의 의미를 확립하는데 도움을 줄 수 있을 것으로 기대하고 있다(Openshaw, 1990; Fischer and Nijkamp, 1992; Goodchild *et al.*, 1992; Bailey, 1994; Fotheringham and Charlton, 1994; Openshaw and Clarke, 1996; Unwin, 1996; Anselin, 1998; Wise *et al.*, 1999). ESDA의 성장과 밀접히 관련되어 있는 또 하나의 요소는 국지적 통계량이다. 공간데이터 분석에서 최근 경향은 연구 대상 지역 전체의 평균적 경향을 기술하거나 모델링하는 것에서 벗어나, 각 하위 지역이 보유한 본질적인 특이성에 주목함으로써 ‘공간적 이질성(spatial heterogeneity)’을 시각화하고 모델링하려는 방향으로 진화하고 있다. 이러한 경향은 소위 ‘국지적 전회(local turn)’라고 불릴 수 있을 만큼 공간데이터 분석의 전 영역으로 확산되고 있다(Fotheringham, 1997; 2000).

이러한 국지적 통계량이 ESDA와 결합하여 가장 활발히 논의되고 있는 영역이 ‘공간 연관성 측도(spatial association measures; 이하 SAM)’와 관련된 분야이다. 전역적 SAM으로부터 도출된 local Moran's  $I_i$ , local Geary's  $c_i$ , Getis-Ord's  $G_i^*$ 와  $G_i^*$ 와 같은 일변량 국지적 SAM은 ‘국지적 공간 연관성 지표(local indicators of spatial association; 이하 LISA)’라고 하는 통계량의 한 클래스를 구성하고 있다(Anselin, 1995; Getis and Ord, 1996). 이러한 일변량 LISA에서 더 나아가 이변량 및 다변량 LISA를 개발하려는 시도 역시 진행되고 있다(Lee, 2001b; Anselin *et al.*, 2002). 이러한 LISA가 ESDA에서 가지는 역할은 명확하다. LISA와 관련 시각화 기법을 사용함으로써 연구자는 공

간적 변동(spatial variation) 속에 국지적으로 존재하는 공간적 클러스터(spatial cluster), 공간적 특이점(spatial outliers), 공간적 체제(spatial regimes)를 확인할 수 있다(Anselin, 1995).

본 연구의 주된 목적은 위에서 살펴 본 국지적 SAM에 기반한 ESDA 방법론이 거주지 분화 연구에 새로운 통찰력을 제시해 줄 수 있다는 점을 논증하는 것이다. 구체적인 연구 과제는 다음과 같다. 첫째, 전역적 공간 분리성 측도로부터 국지적 공간 분리성 측도를 도출하고 유의성 검정법을 제시한다. 둘째, 국지적 공간 분리성 측도를 이용한 이변량 ESDA 기법을 개발한다. 셋째, 개발된 이변량 ESDA 기법을 우리나라 7대 도시의 동별 거주지 분리 양상에 적용한다. 이 사례 연구에서는 이전의 전역적 연구와 마찬가지로 40~59세 인구를 대상으로 고학력 집단과 저학력 집단 간의 공간적 거주지 분리의 양상이 분석된다.

## 2. 국지적 공간 분리성 측도(local spatial separation measure)

### 1) 국지적 공간 분리성 측도의 도출

이상일(2007, 623)은 공간 연관성 측도와 공간적 카이-스퀘어 통계량을 결합하고 이변량 상황을 고려함으로써 이변량 Geary와 형식을 공유하는 전역적 공간 분리성 측도(spatial separation measure; SSM)를 제시하였는데, 그 수식은 다음과 같다.

$$SSM = \frac{n-1}{2\sum_j v_{ij}} \frac{\sum_j v_{ij} [(r_i^X - p_i) - (r_j^X - p_j)] [(r_i^Y - p_i) - (r_j^Y - p_j)]}{\sqrt{\sum_i (r_i^X - p_i)^2} \sqrt{\sum_i (r_i^Y - p_i)^2}} \quad (1)$$

여기서  $v_{ij}$ 는 공간단위 간 위상 관계를 정의하는 공간 근접성 행렬(spatial proximity matrix)의 한 요소이고,  $r_i^X$ 는 특정 인구 집단  $X$ 의 총 인구 중  $i$ 번째 공간단위에서 분포하는 인구의 비중, 즉 열-비중(column-

proportions)을,  $r_j^X$ 는  $i$ 번째 공간단위와 인접한 공간단위에 분포하는  $X$  집단 인구의 열-비중을 의미하고,  $p_i$ 는 모든 인구 집단을 포괄한 전체 인구 중  $i$ 번째 공간단위에서 분포하는 인구의 열-비중을,  $p_j$ 는  $i$ 번째 공간단위와 인접한 공간단위에 분포하는 전체 인구의 열-비중을 의미한다. 또한  $r_i^Y$ 와  $r_j^Y$ 는 또 다른 인구 집단  $Y$ 에 대한 것이다. 여기서 공간적 인접성은 반드시 경계를 접해야 한다는 연접성(contiguity)을 요구하는 것이 아니라, 근접성(proximity)을 규정하는 다양한 방식에 따라 상이하게 규정될 수 있다.

만일 행-표준화된(row-standardized) 공간 근접성 행렬이 적용되었다면, 식(1)은 다음과 같이 훨씬 단순화된 형태로 주어질 수 있다.

$$SSM = \frac{n-1}{2n^2} \cdot \sum_j w_{ij} (z_i^X - z_j^X) (z_i^Y - z_j^Y) \quad (2)$$

여기서  $w_{ij}$ 는 행-표준화된 공간 근접성 행렬의 요소이고,  $z_i^X$ 와  $z_i^Y$ 는 각 공간단위 별 두 인구 집단의 열-비중을 표준화한 값으로, 다음의 수식에 의해 주어진다. 이것은 각 인구집단의 집중도에 대한 표준화 값을 제공하는데, 이에 대해서는 뒤에서 상세히 다루어진다.

$$z_i^X = \frac{r_i^X - p_i}{\sqrt{\sum_i (r_i^X - p_i)^2 / n}} \quad \text{그리고}$$

$$z_i^Y = \frac{r_i^Y - p_i}{\sqrt{\sum_i (r_i^Y - p_i)^2 / n}} \quad (3)$$

이 전역적 측도로부터 가법성 요구(additivity requirement), 즉 전역적 측도는 국지적 측도의 평균과 동일하다는 기준을 만족시키는 국지적 공간 분리성 측도( $SSM_i$ )를 추출하면, 다음과 같은 수식으로 주어진다.

$$SSM_i = \frac{n(n-1)}{2\sum_j v_{ij}} \frac{\sum_j v_{ij} [(r_i^X - p_i) - (r_j^X - p_j)] [(r_i^Y - p_i) - (r_j^Y - p_j)]}{\sqrt{\sum_i (r_i^X - p_i)^2} \sqrt{\sum_i (r_i^Y - p_i)^2}} \quad (4)$$

이 수식을 이차형태(quadratic form)로의 전환을 위해 필수적인 행렬의 형태로 표현하면 다음과 같이 주어진다.

$$SSM_i = \frac{n-1}{2} \cdot \frac{(\mathbf{z}^X)^T [\mathbf{\Omega}_i - (\mathbf{V}_i + \mathbf{V}_i^T)] \mathbf{z}^Y}{\mathbf{1}^T \mathbf{V}_i \mathbf{1}} \quad (5)$$

식(5)에 나타나 있는 국지적 공간 분리성 측도를 위한 공간 인접성 매트릭스는 다음과 같이 다소 복잡한 형태를 취하는데, 이는 유의성 검정을 위해 필수적인 과정이다(Lee, 2008).

$$\mathbf{\Omega}_i - (\mathbf{V}_i + \mathbf{V}_i^T) = \begin{bmatrix} v_{i1} & \mathbf{0} & -v_{i1} & & \\ \mathbf{0} & \ddots & \vdots & \mathbf{0} & \\ -v_{i1} & \cdots & \sum_j v_{ij} - v_{ii} & \cdots & -v_{in} \\ & & \mathbf{0} & \ddots & \mathbf{0} \\ & & -v_{in} & \mathbf{0} & v_{in} \end{bmatrix} \quad (6)$$

여기서,  $\mathbf{\Omega}_i$ 는  $\{v_{i1}, \dots, (v_{ii} + \sum_j v_{ij}), \dots, v_{in}\}$ 으로 이루어진 대각행렬(diagonal matrix)이며,  $\mathbf{V}_i$ 는  $\{v_{i1}, \dots, v_{ii}, \dots, v_{in}\}$ 가  $i$ 번째 행에 놓여지고 나머지 원소는 모두 0인 행렬이고,  $\mathbf{V}_i^T$ 는  $\mathbf{V}_i$ 의 전치행렬이다.

행-표준화된(row-standardized) 공간 근접성 행렬을 가정한 식(2)로부터 국지적 공간 분리성 척도를 도출하면 다음과 같이 보다 단순화된 수식으로 주어진다.

$$SSM_i = \frac{n-1}{2n} \cdot \sum_j w_{ij} (z_i^X - z_j^X) (z_i^Y - z_j^Y) \quad (7)$$

이 수식을 살펴보면, 각 공간단위에 대한 공간 분리성 척도는 이웃 공간단위와의 표준화된 열-비중 상에서의 차이를 두 집단 각각에 대해 구하고 그것들 간의 교차-곱(cross-product)에 대한 가중 평균값으로 산출된다는 사실을 알 수 있다. 예를 들어, 한 공간단위 내에서 한 집단이 다른 집단에 비해 과대표성을 보이고(over-represented), 그러한 양상이 이웃 공간단위에서도 유사하게 나타난다면,  $SSM_i$ 는 0에 근접하게 될 것이다. 전역적 공간 분리성 측도에 대한 해석에서도 보았듯이(이상일, 2007), 척도값이 0에 가까울수록 공

간적 의존성이 높다는 것을 의미한다. 국지적 공간 분리성 측도의 해석에 대해서는 이후에 자세하게 다루어질 것이다.

## 2) 유의성 검정

국지적 통계량은 전역적 통계량과 달리 유의성 검정에 대해 어떠한 가정을 하느냐에 따라 통계량의 분포에 대해 다른 결론에 도달하게 되고, 검정 방법도 달라지게 된다. 국지적 통계량의 통계적 검정을 위한 가정은 크게 두 가지인데, ‘정규성 가정(normality assumption)’과 ‘랜덤화 가정(randomization assumption)’이 그것이다(Cliff and Ord, 1981). 이 때 정규성 가정은 표본이 정규분포를 이루고 있는 모집단으로부터 임의적으로 추출된 것이라고 가정하는 것이고, 랜덤화 가정은 주어진 표본이 발현할 수 있는 모든 공간적 패턴들의 총합을 하나의 초-모집단(super-population)으로 가정한다. 따라서 정규성 가정에서는 측정치들이 표본마다 다르게 구성될 수 있지만, 랜덤화 가정에서는 측정치들은 변하지 않은 채 수치들의 공간적 배열상만 달라지게 된다. 만일 표본의 개수가  $n$ 이고 값이 모두 다르다면, 생성될 수 있는 모든 임의적 순열(random permutation)의 수, 즉 공간적 패턴의 수는  $n!$ 로 주어진다(Lee, 2001a). 이 모든  $n!$ 개의 공간적 패턴에 대해 통계량을 산출하고 기대값과 분산을 구하게 되면 주어진 공간적 패턴이 임의적인 과정을 통해 형성된 것인지 그렇지 않은지를 검정할 수 있게 된다.

이변량 상황에서 랜덤화 가정은 보다 복잡한 개념화를 요구한다. 이변량 상황은 각 공간단위가 측정치의 쌍을 가지고 있다는 것을 의미하는데, 각 쌍을 구성하는 두 값은 임의적 순열을 생성할 때 절대 분리되지 말아야 한다. 이것의 분리를 용인하게 되면 공간통계학적인 의미에서는 중요성을 갖지 않는 표본분포가 도출된다(Lee, 2008). 일변량에서와 마찬가지로 표본의 수가  $n$ 이고 각 쌍 내에서 두 값의 관계가 모두 다르다면, 총  $n!$ 개의 공간 패턴의 쌍을 갖게 된다. 각 공간 패턴의 쌍에 대해 이변량 통계량이 산출되면, 일변량의 경우와 동일한 방식으로 가설검정을 행할 수 있게 된다.

국지적 일변량 SAM에 대한 가설검정에서 이 두 가정은 상호보완적으로 사용되고 있는데(Cliff and Ord, 1981; Sokal *et al.*, 1998; Leung *et al.*, 2003), 불행히도 이변량 SAM에 대해 정규성 가정을 적용시키는 것은 불가능하다(Lee, 2001b; 2008). 본 논문에서 제시된 공간 분리성 측도 역시 이변량 SAM의 형태를 띠고 있기 때문에 유의성 검정도 랜덤화 가정에 의존할 수밖에 없다. 그러나 랜덤화 가정은 모집단의 정규분포를 가정하지 않는다는 측면에서(Cliff and Ord, 1981), 그리고 주어진 공간적 패턴을 모든 가능한 공간적 패턴 중의 하나로 바라보는 것이 개념적으로 이해하기 용이하다는 측면에서 정규성 가정에 비해 우월성을 가질 수 있다.

이러한 랜덤화 가정은 또 다시 두 종류로 나뉘어지는 데, ‘총체적(total)’ 랜덤화 가정과 ‘조건적(conditional)’ 랜덤화 가정이 그것이다(Anselin, 1995; Sokal *et al.*, 1998; Lee, 2001b; 2008). 총체적 랜덤화 가정은 전역적 통계량에 대한 랜덤화 가정과 개념적으로 동일하다. 즉,  $n!$  개의 공간적 패턴의 쌍 각각에 대해 국지적 통계량을 산출하면, 한 공간단위는 총  $n!$  개의 국지적 통계량을 갖게 될 것이고, 그 값들의 표본분포에 대한 정보를 통해 국지적 패턴이 임의적 과정에 의해 생성된 것인지 그렇지 않은지를 통계적으로 검토할 수 있게 된다. 이에 비해 조건적 랜덤화 가정은 특정한 공간단위에는 오로지 최초의 쌍만이 위치한다는 제한조건 하에서 임의적 순열을 생성한다. 즉, 각 공간단위에 주어진 값의 쌍은 그대로 둔 채 이웃하고 있는 공간단위에 부여되는 값만이 임의적으로 배열되는 것이다. 이웃 공간단위에 값이 임의적으로 부여될 때마다 국지적 통계량을 산출함으로써 국지적 패턴에 대한 가설검정을 행하는 방식을 취한다.

이러한 두 가지 서로 다른 랜덤화 가정은 서로 다른 가설 검정 방법을 요구한다. 총체적 랜덤화 가정에 기반한 국지적 공간 분리성 측도에 대한 가설 검정은 전역적 측도에 대한 가설 검정에서와 마찬가지로 ‘확장-만텔 검정(extended Mantel test)’에 의해 수행된다(Lee, 2004c). 식(5)를 이용하면, 국지적 공간 분리성 측도가 전역적 측도와 마찬가지로 두 행렬의 케이스 별 곱의 합으로 표현될 수 있음을 알 수 있다. 유일한

차이는  $\mathbf{P}$ 가  $\mathbf{P}^{(i)}$ 로 대체된다는 사실뿐이다(이상일(2007)의 식(16)과 식(17) 참조).

$$\mathbf{P}^{(i)} \equiv \frac{n-1}{2} \cdot \frac{[\boldsymbol{\Omega}_i - (\mathbf{V}_i + \mathbf{V}_i^T)]}{\mathbf{1}^T \mathbf{V}_i} \quad (8)$$

$$\mathbf{Q} \equiv \mathbf{z}^X \cdot (\mathbf{z}^Y)^T \quad (9)$$

이것을 확장-만텔 검정이 제시하고 있는 틀 속에 투입하면(Lee, 2004c), 총체적 랜덤화 가정하에서의 기대값을 구하는 수식이 다음과 같이 도출된다.

$$E(SSM_i) = \frac{n}{\mathbf{1}^T \mathbf{V}_i} \cdot (v_i - v_{ii}) \cdot r_{XY} \quad (10)$$

여기서  $v_i = \sum_j v_{ij}$ 를 의미한다. 만일 공간 근접성 행렬의 대각선 상에 0이 위치하고, 행렬이 행-표준화되었다면, 식(10)은 간단히  $E(SSM_i) = r_{XY}$ 가 된다. 즉, 기대값은 두 변수  $X$ 와  $Y$ 간의 Pearson 상관계수가 되며, 이 때 변수  $X$ 와  $Y$ 는 두 집단의 표준화된 열-비중을 의미한다(이상일(2007)의 식(20)을 참조할 수 있고, 본 논문의 식(3)을 사용해도 결과는 동일함). 또한 분산을 구하는 식은 다음과 같이 주어진다(도출 과정은 Lee(2008) 참조).

$$\text{Var}(SSM_i) = \frac{n-1}{4n^2} \cdot \left\{ \frac{2n(v_i^{(2)} - 1)(1 - r_{XY}^2)}{n-2} - \frac{(2nr_{XY})^2}{n-1} + \frac{1}{(v_i^{(2)} + 3) \cdot nr_{XY}^2 + (v_i^{(2)} + 1) \cdot nb_1^{XY} \cdot r_{XY}^2 + 2n \cdot v_i^{(2)}} \right\} \quad (11)$$

여기서  $v_i^{(2)} = \sum_{j=1}^n v_{ij}^2$ ,  $b_1^{XY} = m_2^{XY} / (m_1^{XY})^2$ ,  $m_1^{XY} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) / n$ ,  $m_2^{XY} = \sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2 / n$ 를 의미한다.

조건적 랜덤화 가정에 기반한 유의성 검정에는 확장-만텔 검정법이 부적절하다. 왜냐하면 조건적 랜덤화 가정 하에서  $SSM_i$ 는 더 이상 두 행렬의 케이스 별 곱의 합으로 정의되지 않기 때문이다. 대신에  $SSM_i$ 는 두 벡터의 케이스 별 곱의 합으로 정의되며,

이것은 ‘일반화된 벡터 랜덤화 검정(generalized vector randomization test)’이라는 새로운 가설 검정법을 요구한다(Lee, 2001b; 2008). 일반화된 벡터 랜덤화 검정법은 통계량이 두 벡터의 케이스 별 곱의 합으로 정의될 때, 그 통계량의 기대값과 분산을 산출할 수 있는 공식을 제공한다(Hubert, 1984; 1987).

$$E(\Gamma) = \frac{1}{n} \cdot \sum_i p_i \cdot \sum_i q_i \quad (12)$$

$$\text{Var}(\Gamma) = \frac{1}{n-1} \cdot \sum_i (p_i - \bar{p})^2 \sum_i (q_i - \bar{q})^2 \quad (13)$$

문제는  $SSM_i$ 가 두 벡터의 케이스 별 곱의 합으로 정의될 수 있으나인데, 식(5)와 식(6)을 이용하면 다음과 같은 두 벡터를 정의할 수 있게 된다.

$$\mathbf{p}^{(i)} \equiv \frac{n-1}{2 \cdot \mathbf{1}^T \mathbf{V} \mathbf{1}} \times [v_{i1}, \dots, -(v_i - v_{ii}), \dots, v_{in}]^T \quad (14)$$

$$\mathbf{q}^{(i)} \equiv \mathbf{z}^X \mathbf{z}^Y - (z_i^X \cdot \mathbf{z}^Y + z_i^Y \cdot \mathbf{z}^X) \quad (15)$$

식(12)~(15)에 따라 조건적 랜덤화 가정에 기반한  $SSM_i$ 의 기대값에 대한 수식을 도출하면 다음과 같다.

$$E(SSM_i) = \frac{n}{2 \cdot \mathbf{1}^T \mathbf{V} \mathbf{1}} \cdot (v_i - v_{ii}) \cdot (z_i^X \cdot z_i^Y + r_{XY}) \quad (16)$$

총체적 랜덤화 가정에서와 마찬가지로, 만일 공간 근접성 행렬의 대각선 상에 0이 위치하고, 행렬이 행-표준화되었다면, 식(16)은 간단히  $E(SSM_i) = (z_i^X \cdot z_i^Y + r_{XY})/2$ 가 된다. 이 기대값 수식이 의미하는 바는, 조건적 랜덤화 가정에 입각할 경우, 기대값은 전역적인 수준에서의 표준화된 열-비중 간의 상관관계뿐만 아니라 국지적 수준에서의 상관관계에도 영향을 받는다는 사실이다. 만일 국지적 수준에서의 상관관계가 전역적 수준에서의 상관관계와 동일하다면, 기대값은 총체적 랜덤화 가정에서의 기대값과 같아지지만, 나머지 경우에 대해서는 상이한 기대값이 산출된다. 한편 분산은 다음과 같이 좀 더 복잡한 수식에 의해 구해진다

(Lee(2008) 참조).

$$\begin{aligned} \text{Var}(SSM_i) &= \frac{1}{(2n)^2 (n-2)} \cdot [(n-1)(v_i^{(2)} - 1)] \cdot \\ &\quad \left\{ \begin{aligned} &(b_i^{XY} r_{XY} - 2z_{Xi} b_{12}^{XY} - 2z_{Yi} b_{21}^{XY} + \\ &2z_{Xi} z_{Yi}) \cdot nr_{XY} + (z_{Xi}^2 + z_{Yi}^2) \cdot n \\ &- z_{Xi}^2 z_{Yi}^2 \\ &-(nr_{XY} + z_{Xi} z_{Yi})^2 \end{aligned} \right\} \quad (17) \end{aligned}$$

총체적 랜덤화와 조건적 랜덤화 가정 각각에 대해 주어진 수식에 의거해 기대값과 분산이 구해지면, 전역적 공간 분리성 척도에서의 경우와 마찬가지로, ‘정규 근사(normal approximation)’를 통해 유의성 검정이 이루어지게 된다.

거주지 분리 연구에서 국지적 지표를 강조한 연구들은 크게 두 부류로 나눌 수 있는데, 하나는 공간적 격리 지수를 국지화 하는 연구(Maly, 2000; Wong, 2002; 2003; Feitosa *et al.*, 2007)와 국지적 일변량 공간 연관성 측도를 이용하는 연구(최은영, 2004; Brown and Chung, 2006; Chung and Brown, 2007)가 있는데, 이 두 가지 접근법 모두에 대해 이 연구에서 제안하고 있는 국지적 공간 분리성 척도는 장점이 있다. 전자의 접근에 대해, 국지적 공간 분리성 척도는 전역적 공간 분리성 척도와 직접적으로 연결되어 해석될 수 있다는 장점이 있다. 즉, 국지적 공간 분리성 척도의 평균 값은 전역적 공간 분리성 척도와 동일하다. 이와 함께 유의성 검정 방법이 제공되기 때문에 진술의 통계적 신뢰성이 담보될 수 있다. 후자의 접근에 대해서는, 공간적 거주지 분리라는 현상이 기본적으로 이변량 상황을 전제하고 있기 때문에 일변량의 관점에서 탐색적 공간데이터 분석을 실행하는 것은 어떤 의미에서 원시적인 것이다. 이상일(2007, 622~623)이 언급한 것처럼, 공간 분리성 척도는 “한 공간단위 내에서 두 집단의 연관성, 한 집단에 대해 한 공간단위와 이웃 공간단위 간의 연관성, 또 다른 집단에 대해 한 공간단위와 그 이웃 공간단위 간의 연관성, 그리고 이웃 공간단위들 내에서의 두 집단의 연관성” 등 네 가지 연관성을 동시에 고려하는 접근법이 보다 진보된 것이다.

### 3) 결과의 해석

이상일(2007)은 전역적 공간 분리성 척도를 해석하기 위해 기대값(혹은 척도값)의 부호와 유의성 검정에서의 z-값의 부호를 결합하는 방식을 제안했다. 그것을 원용하면서 국지적 패턴의 해석이라는 상황에 맞추어 결과 해석의 원칙을 정리하면 다음과 같다(이상일(2007)의 그림 3 참조).

첫째, 전역적 분석에서 기대값이 0보다 작으면 거주지 분리가, 기대값이 0보다 크면 거주지 중첩이 연구대상 지역 전체에서 평균적으로 나타난다고 규정한 바 있다(이상일, 2007, 625). 그리고 그 기대값은 표준화된 열-비중 간의 피어슨 상관계수로 주어진다. 국지적인 차원에서 이러한 기대값을 고려하기 위해 일종의 국지적 피어슨 상관계수를 생각해볼 수 있는데, 이것은 각 공간단위 내에서의 표준화된 열-비중 간의 곱(식(3) 참조)으로 나타난다. 이 때 전역적 피어슨 상관계수는 국지적 피어슨 상관계수의 평균과 동일하다. 각 공간단위에서 이 국지적 피어슨 상관계수가 양수로 주어지면 그 공간단위에서 거주지 중첩이 발생하는 것으로 간주하고, 음수로 주어지면 거주지 분리가 발생하는 것으로 간주한다. 둘째, 전역적 분석에서, 공간 분리성 척도와 기대값과의 관계가 공간적 의존성의 정도를 나타내며, z-값의 절대치가 공간적 의존성의 강도를 표현한다고 해석되었다(이상일, 2007, 626). 이것을 그대로 국지적 공간 연관성 척도의 해석에 적용하면, 국지적 “공간 분리성 척도가 기대값으로부터 0의 방향으로 향하면 양(+)적인 공간적 의존성이 나타나는 것이고, 반대로 기대값으로부터 0과 반대 방향으로 향하면 음(-)적인 공간적 의존성이 나타나는 것을 의미”한다(이상일, 2007, 625~626). 이 때 양(+)적인 공간적 의존성에 대해 ‘클러스터’란 단어를 사용했는데, ESDA 방법론을 위해 뒤에서 제시될 ‘공간적 클러스터’라는 개념과 혼동을 일으키기 때문에 본 논문에서는 ‘집중’이라는 단어를 사용하고자 한다.

이 두 가지 기준을 결합하면 전역적 공간 분리성 척도에서와 마찬가지로 네 가지 경우가 도출된다. 첫째, 기대값(국지적 피어슨 상관계수)이 양수이면서 z-값이 양수이면(중첩-혼재), 두 집단이 해당 공간단위에서

중첩되는 경향이 있으면서 동시에 음(-)적인 공간적 의존성이 나타난다는 것을 의미한다. 둘째, 기대값이 양수이면서 z-값이 음수이면(중첩-집중), 두 집단의 거주 공간이 해당 공간단위 내에서 중첩되는 경향이 있으면서 동시에 양(+)적인 공간적 의존성이 나타난다는 것을 의미한다. 셋째, 기대값이 음수이면서 z-값이 양수인 경우(분리-집중)는 해당 공간단위 내에서 거주지 분리의 경향이 존재하면서 동시에 양(+)적인 공간적 의존성이 나타난다는 것을 의미한다. 전역적 분석에서와 마찬가지로 이것이 공간적 분리 현상의 전형적인 모습이라고 할 수 있다. 넷째, 기대값이 음수이면서 z-값이 음수인 경우(분리-혼재)는 해당 공간단위 내에서는 거주지 분리의 경향이 나타나지만 음(-)적인 공간적 의존성이 지배적이라는 것을 의미한다. 이러한 네 가지 구분법은 뒤에서 제시될 탐색적 공간데이터 분석에서 핵심적인 역할을 담당할 것이고 뒤에서 상술된다.

## 3. 연구 방법론의 정립

### 1) 학력 집단 별 집중도

학력 집단 별 집중도를 표현하는 방식에는 여러 가지가 있을 수 있다. 첫 번째는 가장 단순하면서 가장 널리 쓰이는 방법으로 각 공간단위 내에서 전체 인구 중 특정 학력 집단에 포함되는 인구 수의 비중(행-비중)을 표시하는 방법이다. 두 번째 방법은 첫 번째 방식에 기반하지만 과대표성을 띠는 지역을 추출하기 위해 일종의 표준화를 행하는 방법이다. 여기에는 행-비중을 표준화하는 방법과 입지계수(location quotient)를 계산하는 방법이 있다. 이 두 가지 표준화 방식의 차이는 표준화의 기준점으로 행-비중의 평균을 사용하느냐 아니면 전체 지역에서 특정 인구 집단이 차지하는 비중을 사용하느냐이다. 개념적인 측면에서 입지계수를 사용하는 것이 보다 타당하다고 볼 수 있다(거주지 분리 연구에 입지계수가 사용된 예를 위해 다음의 논문을 참고할 수 있다. Brown and Chung, 2006; Chung and Brown, 2007). 그러나 이러한 접근법은

표 1. 인구 집단 별 집중도 계산의 다양성

공간단위	특정집단 인구 수 <sup>a</sup>	전체 인구 수 <sup>b</sup>	행-비중 (a/b)	특정집단 열-비중	인구전체 열-비중	입지계수	표준화 집중도*
A	40	200	0.2	0.1	0.2	0.5	-0.1
B	240	300	0.8	0.6	0.3	2.0	0.3
C	0	200	0.0	0.0	0.2	0.0	-0.2
D	80	100	0.8	0.2	0.1	2.0	0.1
E	40	200	0.2	0.1	0.2	0.5	-0.1
합계	400	1,000		1.0	1.0		

\* 단순화하기 위해 식(3)에서 분자 부분만 계산함.

통계학적인 의미에서 모두 동일한 한계점을 갖고 있는데, 모두 공간단위의 상이한 표본 개수를 고려하지 않는 오류를 범하고 있다(Oden, 1995; Rogerson, 1999). 즉, 행-비중이 총 인구 수가 매우 큰 공간단위에서 측정된 것인지 매우 작은 공간단위에서 측정된 것인지 구분하지 않는다. 비록 동일한 행-비중이라 하더라도 각 공간단위의 총 인구 규모가 크다면 그 비중값이 갖는 통계적 유의성은 더 커진다. 또한 행-비중이 동일하다면 당연히 입지계수도 동일한 값을 취하기 때문에 이 문제는 입지계수에서도 동일하게 드러난다.

이러한 문제를 해결하기 위해서는 과대표성의 확인을 위한 특정한 표준화 방식을 포함하면서도 공간단위의 상이한 규모 수준을 고려할 수 있는 지표를 산출할 필요가 있다. 여기서는 식(3)에 나타난 수식을 제안하고자 한다. 이상일(2007)이 언급한 것처럼, “거주지 분리란 기본적으로 공간단위라고 하는 범주를 가로질러 인구 집단이 어떤 빈도 분포를 보이고 있는냐의 문제를 의미하며, 그 빈도 분포가 이질적일 때 우리는 두 집단 간의 거주지 분리가 존재한다고 말할 수 있기 때문”(621)에, “각 공간단위의 특정 인구 집단의 열-비중을 전체 인구 집단에서 각 공간단위가 차지하는 비중으로 표준화하는 것은 매우 합리적”(622)이다. 식(3)을 살펴보면, 전체 인구에서 각 공간단위 전체 인구의 비중이 특정한 인구 집단 전체에서 각 공간단위 내의 그 집단 인구의 비중을 위한 기대값으로 사용되고 있음을 알 수 있다. 표 1은 이러한 논리를 가상의 자료를 통해 도해하고 있다. 한 도시가 A-E라는 다섯 개의 공간단

위로 구성되어 있고, 도시 전체 인구는 1,000명, 해당 인구 집단의 총 인구는 400명이라고 할 때, 각 공간단위 별로 집중도 산출과 관련된 통계량은 표 1과 같이 주어질 수 있다. 이 때 B와 D의 대조는 중요하다. 행-비중과 입지계수는 두 공간단위 모두에 대해 0.8과 2.0의 값을 가진다. 그러나 B와 D의 전체 인구는 각각 300명과 100명이기 때문에 통계적 유의성이라는 측면에서 동일하게 취급될 수 없다. 따라서 마지막 열에서 보는 것처럼, 식(3)에 따른 표준화 집중도가 이러한 차이를 잘 반영할 수 있다.

## 2) 국지적 공간 분리성 측도를 이용한 ESDA 기법

국지적 공간 분리성 측도의 해석에서 제시된 네 가지 구분법을 이용하면, 거주지 분리 현상에 대한 ESDA 기법이 도출될 수 있다. 이를 위해 두 가지 개념에 대한 논의가 필요한데, ‘공간적 체제(spatial regimes)’와 ‘공간적 이례치(spatial anomalies)’가 그것이다. 이변량 상황에서의 공간적 체제란 두 변수 간의 교차 비교를 통해 도출되는 범주 설정에서 동일한 범주에 포함되는 공간단위의 연속체이다. 이것은 일변량의 경우 Anselin(1996)이 제안한 ‘Moran 산포도 지도(Moran scatterplot map)’와 유사한 것이다. 이 논문에서 제안된 공간 분리성 측도의 기대값과 z-값을 교차시키면, 중첩-혼재, 중첩-집중, 분리-집중, 분리-혼재의 네 범주가 도출되고, 이것이 일종의 명목 지도로 표현되었을 때 각 범주에 포함되는 공간단위의 연속체가 하

나의 공간적 체제로 인식될 수 있다. 본 연구의 맥락에서 보면, 중첩에 해당되는 첫 번째 두 범주에 속한다는 것은 고학력 집단의 표준 점수와 저학력 집단의 표준 점수의 부호가 각 공간단위 내에서 동일하다는 것을 의미하고, 분리에 해당하는 나머지 두 개의 범주는 그 부호가 반대라는 것을 의미한다. 따라서 중첩과 분리의 구분은 공간적 의존성은 배제한 채 공간단위 별 분리/중첩의 상황만을 고려하는 비공간적 방식이다.

이와는 대조적으로 혼재/집중의 구분은 공간단위 별 분리/중첩이 공간적 의존성을 보이는지 그렇지 않은지를 말해준다. 혼재는 중심 공간단위와 주변 공간단위들이 보여주는 분리/중첩의 양상에 이질성이 존재하는 경우(음(-)적 공간적 의존성)를 말하는데, 이는 두 가지 이유로 발생한다. 하나는 분리/중첩의 양상과 정도가 주변 공간단위 간에는 유사한데, 중심 공간단위와는 다르게 나타나는 경우이고, 또 다른 하나는 중심 공간단위의 분리/중첩 경향과는 관계 없이 주변 공간단위 간에 그러한 경향에 일체성이 결여된 경우에 발생한다. 집중은 중심 공간단위와 주변 공간단위들이 보여주는 분리/중첩의 양상에 동질성이 존재하는 경우(양(+)적 공간적 의존성)를 의미하는 데, 이것 역시 두 가지 이유로 발생한다. 하나는 주변 공간단위 간에 분리/중첩 경향에서 다소간의 이질성이 존재한다 하더라도 평균적인 경향이 중심 공간단위의 분리/중첩 경향과 유사한 경우이고, 또 다른 하나는 중심 공간단위의 분리/중첩 경향과 관계 없이 주변 공간단위 간에 그러한 경향의 일체성이 현저한 경우에 발생한다. 이러한 해석 상의 복잡성은 이 연구에서 제시된 통계량의 한계에 기인한 것인데, 결론 부분에서 상세히 다루어질 것이다. 어쨌든 이렇게 네 개의 범주로 공간단위를 분류하고 그것을 통해 제작된 일종의 명목 지도를 여기서는 ‘공간 분리성 산포도 지도(spatial separation scatterplot map; 이하 SSSM)’ 라고 부르고자 한다.

공간적 이례치란 통계적으로 유의한 만큼의 특이성을 보이는 공간단위를 의미한다. 공간적 이례치는 다시 두 개의 카테고리로 구분되는데, 공간적 클러스터(clusters)와 공간적 특이점(outliers)이다. 공간적 클러스터는 통계적 유의성을 가질 만큼 정(+)적인 공간적 의존성이 현저하게 드러나는 공간단위 혹은 그것들의

연속체를 의미하고, 반대로 공간적 특이점은 통계적으로 유의성을 가질 만큼 부(-)적인 공간적 의존성이 현저하게 드러나는 공간단위를 의미한다. 공간적 이례치의 확인은 위에서 제시된 SSSM으로부터 통계적 유의성을 가진 공간단위 만을 선별함으로써 가능해진다. 즉, 중첩-집중과 분리-집중 범주에 속한 공간단위 중 유의성 검정을 통과하면 공간적 클러스터인 것으로 인정될 수 있고, 중첩-혼재와 분리-혼재에 속한 공간단위 중 유의성 검정을 통과하면 공간적 특이점인 것으로 인정될 수 있다. 통계적 검정 절차를 통해 도출된 이러한 지도를 여기서는 ‘공간 분리성 이례치 지도(spatial separation anomalies map; 이하 SSAM)’ 라고 부르고자 한다.

SSAM에서 공간적 클러스터로 인정된 중첩-집중은 각 공간단위 내에서 두 집단 간 거주지 중첩이 상당한 정도 일어나면서 동시에 그 주변지역의 동들에도 그러한 경향이 현저한 것을 의미하고(혹은 주변지역의 거주지 중첩/분리 경향이 상당히 동질적으로 나타나는 경우), 분리-집중은 각 공간단위 내에서 두 집단 간 거주지 분리가 일어나면서 동시에 그 주변지역도 그러한 경향이 현저한 경우로(혹은 주변지역의 거주지 중첩/분리 경향이 상당히 동질적으로 나타나는 경우), 공간적 거주지 분리 현상이 핵심적으로 드러나는 공간 클러스터를 의미한다. 이와 유사하게, 통계적으로 유의한 중첩-혼재 지역은 각 공간단위 내에서는 거주지 중첩이 상당한 정도 일어나지만 그 주변 지역에서는 거주지 중첩과 분리가 혼재되어 나타나는 지역을 의미하고(혹은 중심 공간단위와 주변 공간단위 간의 일치도가 현저히 떨어지는 경우), 분리-혼재 지역은 각 공간단위 내에서는 거주지 분리가 상당한 정도 일어나지만 그 주변 지역에서는 거주지 중첩과 분리가 혼재되어 나타나는(혹은 중심 공간단위와 주변 공간단위 간의 일치도가 현저히 떨어지는 경우) 경우를 의미한다.

#### 4. 한국 대도시에서의 적용

이 연구에서 사용된 데이터는 전역적 연구(이상일,

2007)에서와 마찬가지로, 2000년 인구 센서스 원자료 CD에서 추출한 40~59세 인구의 최종학력 자료이다. 최종학력을 고학력(4년제 대졸 이상), 중학력(중졸 이상에서 4년제 대졸 미만), 저학력(중졸 미만)으로 구분하였는데, 본 국지적 연구에서는 고학력과 저학력 집단 사이의 거주지 분리만 다룬다. 모든 계산은 S-Plus가 제공하는 S-언어를 사용하여 저자가 작성한 스크립트에 의해 이루어졌다.

각 도시에 대한 국지적 분석을 위해 위의 연구방법론에서 제시된 기법이 사용되었다. 각 도시에 대해 우선적으로 학력 집단별 집중도 분포를 살펴본 후, SSSM과 SSAM을 통해 ESDA가 시행된다. 국지적 공간 분리성 측도에 대한 유의성 검정은 조건적 랜덤화 가정에 기반했고, SSAM 작성을 위한 유의수준으로 0.10이 선정되었다. 조건적 랜덤화가 사용된 것은 총체적 랜덤화에 비해 패턴 확인에 유리하기 때문이다(Anselin, 1995), 유의수준으로 전통적인 0.05나 0.01을 사용하지 않은 것은 패턴의 탐색이 ESDA의 가장 중요한 목적이기 때문에 해당 연구에 따라 다양한 유의수준이 설정될 수 있기 때문이다(Tiefelsdorf, 1998; Sokal *et al.*, 1998; Lee, 2004c).

## 1) 서울

고학력 집단에서 과대표성을 보이는 동(2.0 이상)은 총 33개 동인데, 이 중 23개가 강남구, 서초구, 송파구 등 소위 강남 3구에서 집중 분포하고 있다(그림 1 참조). 상위 10개 동은 강남구 대치2동(5.59), 송파구 오륜동(4.81), 강남구 대치1동(4.65), 영등포구 여의도동(4.49), 양천구 목6동(4.19), 서초구 잠원동(4.00), 서초구 서초4동(3.97), 양천구 목5동(3.46), 서초구 반포3동(3.29), 강남구 압구정1동(3.26) 등이다. 저학력 집단이 과대표성을 보이는 동(2.0 이상)은 모두 5개로 나타났다.

SSSM 지도를 보면, 매우 높은 전역적 수준의 거주지 분리 경향 때문에(전역적 기대값  $-0.8647$ ), 총 522개 동 중 451개 동이 분리 범주에 속한다. 분리 범주에 속하는 동 중 분리-집중에는 389개, 분리-혼재에는 62개가 속한다. SSSM에 유의성 검정을 적용시켜 제작된

SSAM 지도를 보면, 총 106개 동이 선별되었는데, 이 중 71개 동이 분리-집중 범주에, 62개 동이 분리-혼재 범주에 포함된다. 지도에서 보는 바처럼 강남을 중심으로 거대한 거주지 분리의 공간 클러스터가 나타나고 있다. 이 클러스터의 핵심에 강남구의 대치1동, 대치2동, 삼성2동과, 서초구의 서초4동과 반포4동이 위치해 있다. 분리-혼재에 해당하는 동 중 대표적인 것이 용산구 보광동인데, 그 주변의 동들 중 한강 이남에 위치한 서초구 반포3동과 잠원동은 매우 높은 수준의 거주지 분리 경향을 보이는 반면, 용산구 이태원1동과 한남2동은 거주지 중첩의 경향을 보이고 있고, 용산구 서빙고동과 한남1동은 낮은 수준의 분리 경향을 보여주는 등 주변 동의 분리/중첩 경향에 일차성이 심대하게 결여되어 있다. 분리-혼재 범주에 속하는 다른 동도 이처럼 분리-집중 주변부나 분리-집중 지역 내에 고립적인 섬처럼 분포하고 있다.

중첩-혼재 범주에 속하면서 통계적인 유의성을 보이는 유일한 동이 송파구 삼전동인데, 주변의 강남구 일원2동과 송파구의 석촌동과 가락1동이 상대적으로 높은 거주지 중첩의 경향을 보여주고 있는 반면, 송파구의 잠실3동과 잠실본동은 거주지 분리의 경향을 강하게 보여주고 있다. 중첩-집중을 보여주고 있는 동들 중 대표적인 것이 송파구 방이2동인데, 주변의 동들은 모두 거주지 분리의 경향을 강하게 보여주고 있다. 따라서 송파구 방이2동은 거주지 분리가 극심한 지역에 마치 고립된 섬처럼 존재하면서 거주지 중첩 경향을 보이는 공간적 특이점(outlier)의 전형적인 예이다. 중첩-집중에 포함되는 나머지 동들도 모두 자신은 상당한 정도의 거주지 중첩을 보이지만 주변의 거주지 분리 경향이 강한 동들도 둘러싸여 있는 특징을 보여주고 있다.

## 2) 부산

부산의 경우(그림 2 참조), 고학력 집단에서 과대표성을 보이는 동(2.0 이상)은 모두 6개인데, 해운대구 좌동(11.07), 수영구 남천2동(3.88), 금정구 구서2동(3.37), 연제구 연산9동(2.49), 해운대구 우1동(2.37), 동래구 온천2동(2.09) 등이 이에 속한다. 서울과 비교

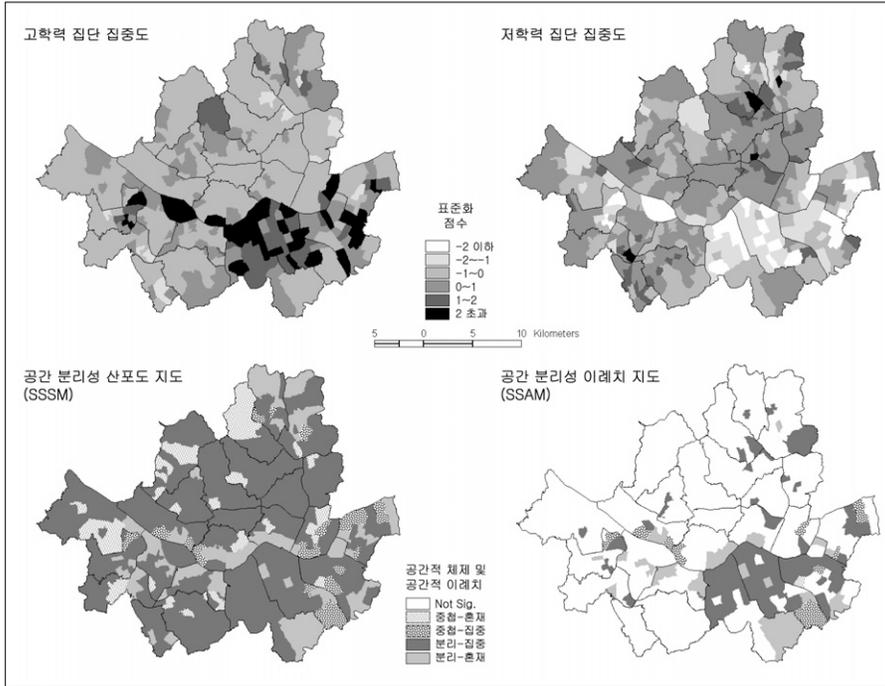


그림 1. 서울의 국지적 거주지 분리 양상

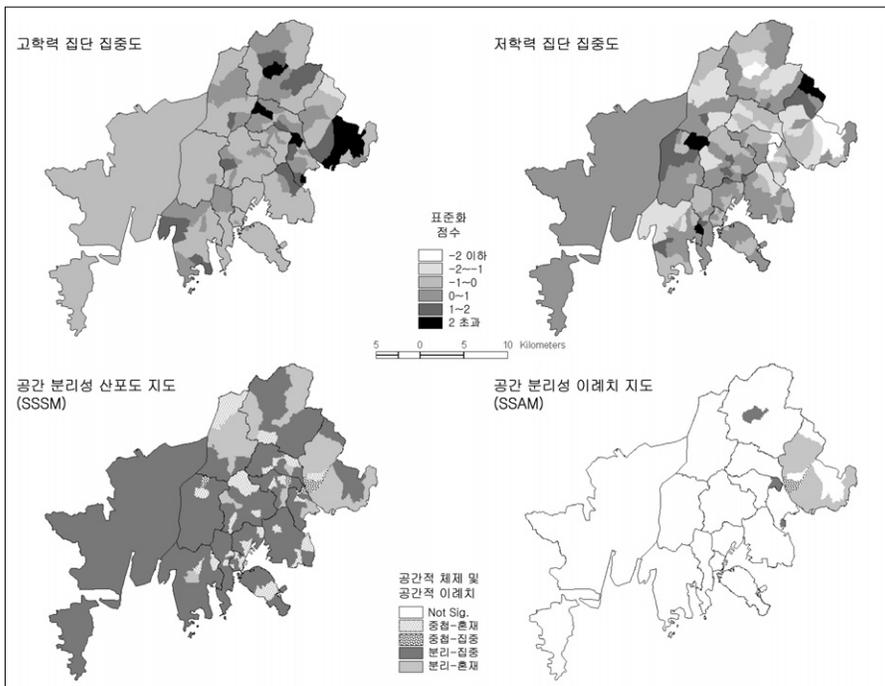


그림 2. 부산의 국지적 거주지 분리 양상

할 때 집중도가 높은 동의 숫자가 훨씬 적을 뿐만 아니라, 높은 집중도를 보이는 동들이 서로 서로 인접해 있기 보다는 여러 구에 흩어져 산포하는 경향이 강하다. 이는 전역적 분석의 결과에도 잘 드러나 있는데(이상일, 2007), 7대 대도시 중 부산만이 고학력-저학력 간에서 분리-혼재, 즉 두 집단간 거주지 분리 경향은 존재하지만, 그러한 경향에서 공간적 의존성은 뚜렷이 드러나지 않는 독특한 양상을 보여주었다. 해운대구 좌동은 7대 대도시 전체에서도 가장 높은 고학력 집단 집중도를 보여주고 있는데, 동 내에서 고학력 집단이 차지하는 비중도 높을 뿐만 아니라(36.4%로 부산에서 2위), 인구 규모도 크기 때문에 매우 높은 값을 나타내었다. 부산에서 저학력 집단의 과대표성을 보이는 동역시 특정한 공간적 클러스터의 형성 없이 산재해 있는 경향이 강하다

SSSM 지도는 분리-집중의 공간적 체제가 전 시역을 장악하고 있는 모습을 보이고 있는데, 이는 전역적 공간 분리성 측도의 기대값(-0.8759)을 그대로 반영하고 있다. 부산은 전역적인 의미에서 분리-혼재 양상을 보이는데 이것은 다른 대도시들이 분리-집중의 경향을 보이는 것과 비교하면 매우 예외적인 것이다. 이러한 부산시의 예외성은 SSAM 지도에 그대로 반영되어 있다. 일단 통계적 유의성을 보인 지역이 전체 216개 동 중 14개에 불과하고, 그 중에서도 분리-혼재를 보이는 동이 10개로 가장 많다. 연제구 연산9동, 금정구 구서2동, 수영구 남천2동이 분리-집중인 것으로 확인되었지만 공간적 확장은 별로 없이 마치 고립된 섬처럼 존재한다. 이것은 그들 동 주변에는 분리의 경향이 강한 동이 존재하지만 그 주변 동 바깥의 동에서는 그러한 경향이 이어지지 않는다는 것을 의미한다. 이것은 지형적 영향으로 인한 선형적인 도시 발달, 내적 동질성이 결여된 공간단위의 설정, 이질적인 공간단위가 행정적으로 연결할 가능성이 높은 점 등에 영향 받은 것으로 보인다.

분리-혼재를 보여주고 있는 10개 동들은 모두 해운대구에 위치하고 있는데, 가장 전형적인 특징을 보이는 것이 반여2동이다. 이 동은 자체적으로 거주지 분리 경향을 보이고 있는데, 주변에 유사한 특성을 보이는 반여1동과 반여3동이 인접해 있음과 동시에 강한 거주

지 중첩 경향을 보이는 제송2동과 제송1동과도 연결해 있다. 해운대구 제송1동은 중첩-집중 범주에 속하는 유일한 동인데 주변에 유사한 특성을 보이는 해운대구 제송2동과 수영구 망미2동이 동 경계를 접하고 있기 때문이다.

### 3) 대구

대구의 경우(그림 3 참조), 고학력 집단에서 과대표성을 보이는 동들이 수성구 내에서 거대한 클러스터를 이루고 있음을 볼 수 있는데, 이러한 경향은 남구를 거쳐 달서구로 이어지고 있다. 상위 10개 중 수성구에 8개, 달서구에 2개가 존재하고 있다. 2.0이 넘는 값을 보인 동들은 모두 6개 인데, 수성구 지산2동(3.75), 고산1동(3.64), 범물2동(3.50), 달서구 상인1동(3.09), 수성구의 범어4동(2.90)과 지산1동(2.75) 등이다. 저소득층에서 과대표성을 보이는 지역은 주로 서구의 전통적인 중심지에 집중되어 있는 경향을 보이는데, 상위 10개 구 중 6개가 남구에 위치하고 있다.

SSSM을 살펴보면, 도시 전역에서 분리-집중 체제가 지배적인 것이 관찰되고(전역적 공간 분리성 측도의 기대값이 -0.8652), 주로 도심 가까운 지역들에서 중첩-혼재 지역이 산포하고 있으며, 특히 남부의 외곽지역에서 중첩-집중 지역이 나타나고 있다. 유의성 기준을 통과한 동은 모두 35개로 이중 분리-집중이 28개로 압도적인 비중을 차지하고 있다. SSAM에서 가장 두드러진 특징은 수성구의 대부분 지역이 거대한 분리-집중의 클러스터를 이루고 있다는 점이다. 특히 수성구의 황금1동, 지산1동, 범어4동, 지산2동, 고산2동, 만촌3동, 범물2동 등은 이 클러스터의 핵심부에 위치한다. 분리-혼재의 전형적인 특징을 보여주고 있는 달서구 상인3동은 달서구 상인1동과 도원동처럼 높은 분리 경향을 보이는 동들과 연결해 있기도 하지만, 중첩의 경향을 보이는 달서구 상인2동과 송현1동과도 경계를 접하고 있다.

### 4) 인천

인천시의 경우(그림 4 참조), 고학력 집단이 과대표

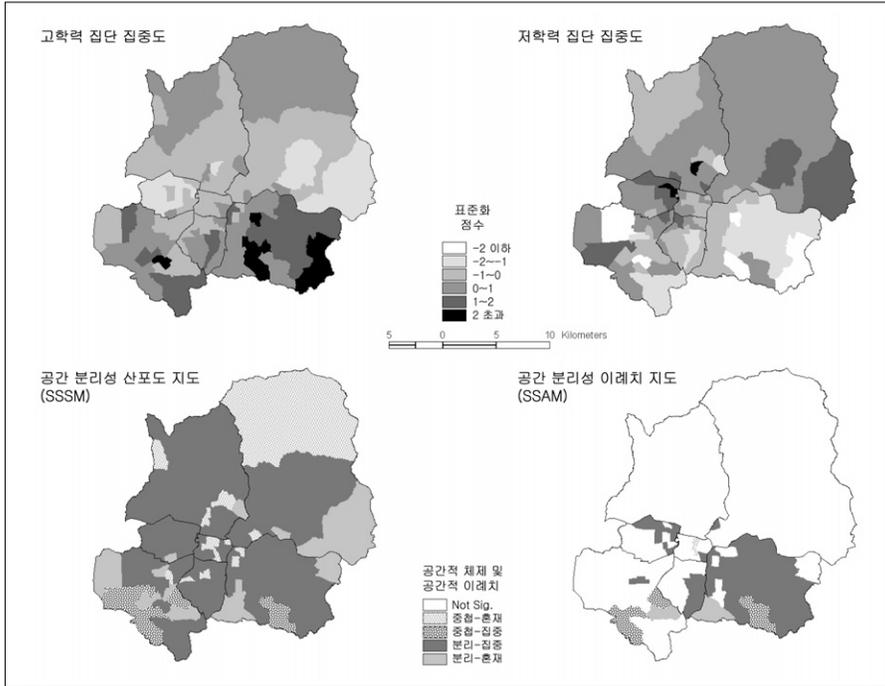


그림 3. 대구의 국지적 거주지 분리 양상

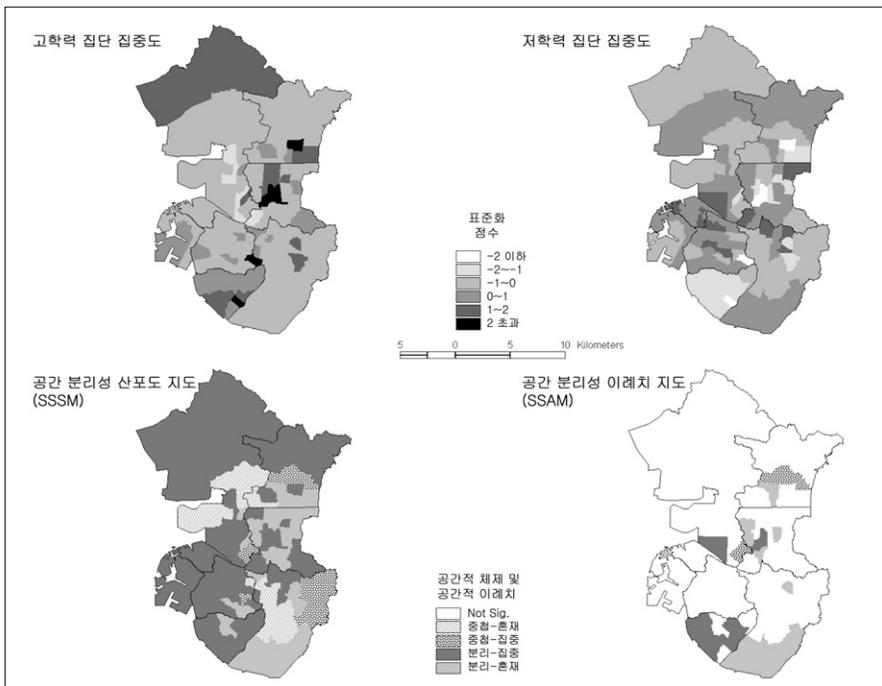


그림 4. 인천의 국지적 거주지 분리 양상

성을 보이는 지역이 소규모의 클러스터 형태로 산재되어 있는 경향이 강하게 나타난다. 2.0 이상을 값을 보이는 동은 모두 6개 인데, 부평구 산곡3동(4.69), 연수구 청량동(4.35)이 월등히 높은 값을 보여주고 있고, 뒤를 이어 부평구의 부평1동(3.27)과 산곡4동(2.77), 계양구 계산3동(2.76), 남구 관교동(2.10) 등이 높은 집중도를 나타내고 있다. 인천시의 고학력 집단의 집중도 패턴에서 두드러지는 것은 상당한 정도의 공간적 의존성을 보이는 고학력 클러스터가 도시 내에 산재되어 있는 형태를 띠고 있다는 것이다. 인천에서는 2.0 이상의 저학력 집단 집중지는 발견되지 않았고, 1~2 사이의 높은 값을 보이는 동들이 특정한 클러스터의 형성 없이 여러 구에 산재하는 경향을 보여주고 있다.

SSSM을 통해 확인할 수 있는 인천시의 공간적 체계는 다른 도시들과 마찬가지로 분리-집중이 압도적인 경향을 보이고, 분리-혼재 지역이 도시의 동부 지역에 남북 방향으로 길게 형성되어 있는 것이 특징적이다. 총 114개 동 중 SSAM에 나타난 동은 총 19개인데, 이 중 9개가 분리-집중, 7개가 분리-혼재에 속한다. 분리-집중이 공간적 클러스터를 이루고 있는 대표적인 지역은 연수구의 연수1동, 청량동, 연수3동, 연수2동, 동춘1동이 연결된 지역이다. 이외에 부평구 산곡3동과 산곡4동이 연결된 지역이 하나의 공간 클러스터로 확인된다. 분리-혼재의 전형적인 특징을 보이는 동이 계양구 계산1동인데, 주변에 높은 분리의 경향을 보이는 계산3동과 같은 동과도 연결해 있지만 중첩의 경향을 보이는 계산2동과 서구의 연희동과도 연결되어 있다. 서구 가좌4동은 중첩-집중의 전형적인 특성을 보이는데, 특히 주변지역들이 거의 비슷한 수준의 분리 경향을 보여주고 있다.

## 5) 광주

광주의 경우(그림 5 참조), 고학력 집단이 과대표성을 보이는 지역이 전통적인 중심지로부터 반경 5~10km 지역에 환형으로 분포하고 있는 특징을 보여주고 있다. 패턴 상에서 보면 서구, 남구, 동구 각각에 클러스터가 형성되어 있는 것처럼 보이지만, 서구를 제외하고는 면적이 큰 동 하나씩이 고립적으로 분포하

고 있다. 이것은 광주시가 전역적 분석 결과 울산광역시와 함께 유의한 공간적 거주지 분리가 나타나지 않았던 점을 잘 반영하는 것이다. 2.0 이상의 집중도를 보인 동은 모두 7개 인데, 남구 봉선2동(4.67)이 탁월하게 높고, 그 뒤를 이어, 서구 상무1동(2.53), 동구 학운동(2.47), 서구의 서창동(2.29)과 화정4동(2.13), 남구 효덕동(2.08), 광산구 비아동(2.00) 등이 있다. 인천과 마찬가지로 2.0 이상의 집중도를 보유한 저학력 집단 집중지는 발견되지 않았다. 상대적으로 높은 값을 보인 동들은 주로 고학력 집단의 집중지 내외곽 지역에 고립 분산적으로 분포하고 있다.

SSSM을 기반으로 광주의 공간적 체계를 살펴보면, 전통적인 중심지로부터 서쪽으로 이어지는 지역에 광범위한 분리-집중 지역이 나타나고, 북부에서 동부를 거쳐 남부로 이어지는 환상의 외곽지역에 분리-혼재 지역이 비교적 넓게 분포하고 있다. SSAM에서 유의성을 가진 동으로 확인된 것은 모두 13개인데, 이 중 분리-집중이 3개, 분리-혼재가 9개, 중첩-집중이 1개 나타나고 있다. 분리-혼재가 분리-집중 보다 많이 확인된 것은 부산의 경우와 동일하다. 분리-집중이 나타나는 동은 광산구 어룡동, 북구 용봉동, 서구 화정2동인데, 모두 고립분산적으로 분포하고 있다. 이 동들은 모두 유사한 특성을 보이는 동들로 둘러 쌓여 있다. 남구 봉선1동은 분리-혼재의 전형적인 특성을 보이고 있는데, 주변의 동에 주월1동처럼 중첩이 나타나는 동이 있는가 하면, 남구 봉선2동처럼 매우 높은 수준의 분리가 나타나는 동도 존재한다.

## 6) 대전

대전의 경우(그림 6 참조), 고학력 집단 집중지가 서구의 클러스터에서 유성구의 보다 규모가 큰 클러스터로 연결되어 있는 것이 확연히 관찰된다. 2.0 이상의 값을 보인 동 6개 모두가 유성구와 서구에 위치해 있다. 그 동들은 유성구의 전민동(4.18)과 신성동(3.42), 서구의 둔산1동(2.91)과 삼천동(2.89), 유성구 온천2동(2.48), 서구 둔산2동(2.01)이다. 인천이나 광주의 경우와 마찬가지로 2.0 이상의 값을 보이는 저학력 집단의 집중지는 대전에서 발견되지 않는데, 주로 도시의 동

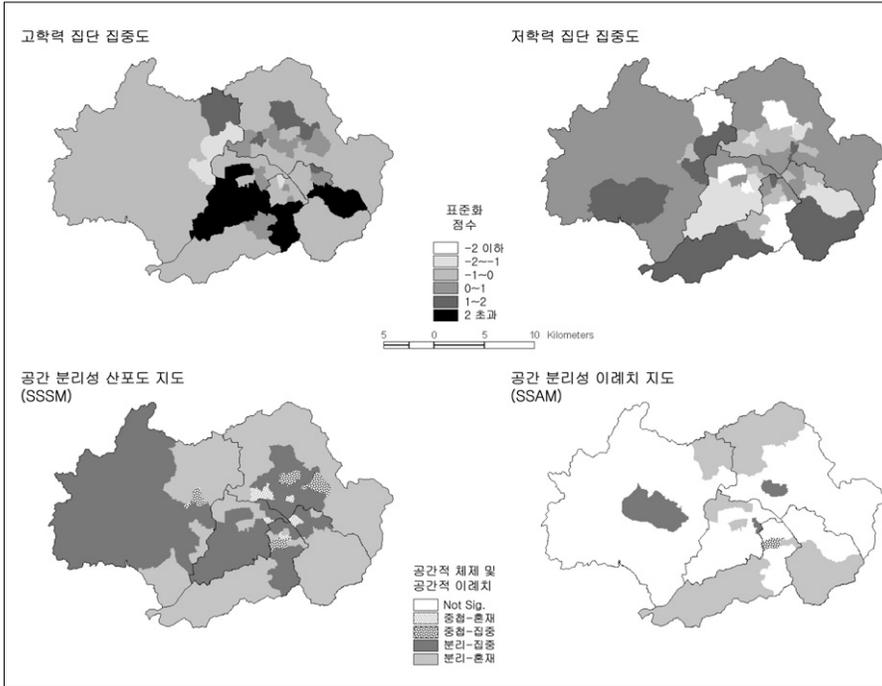


그림 5. 광주의 국지적 거주지 분리 양상

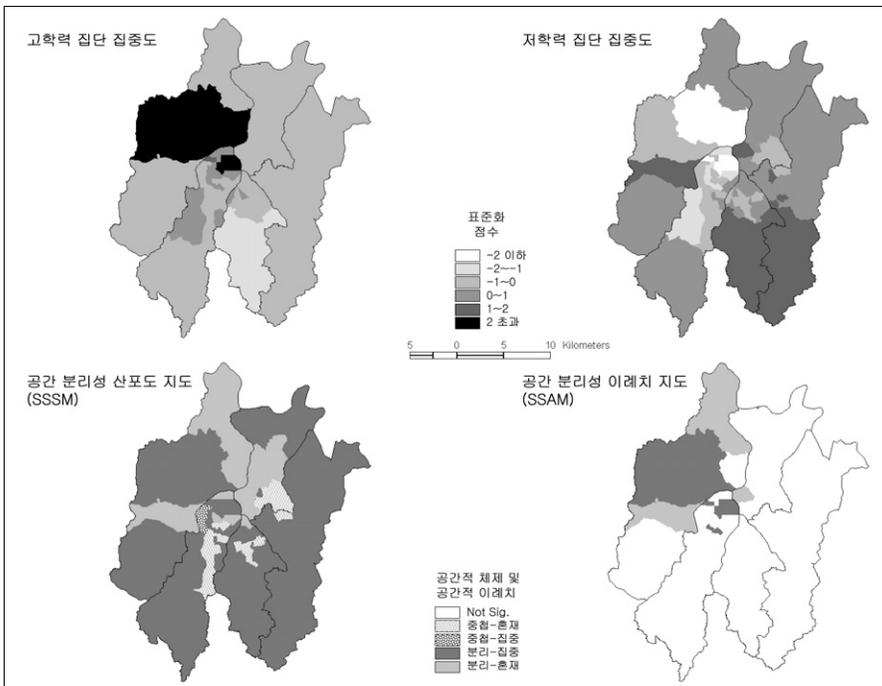


그림 6. 대전의 국지적 거주지 분리 양상

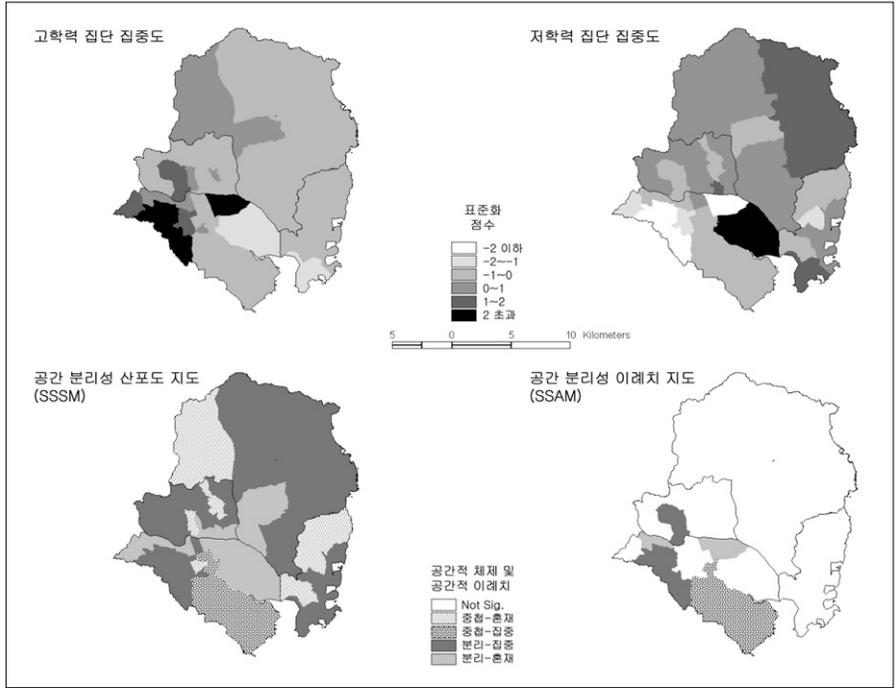


그림 7. 울산의 국지적 거주지 분리 양상

남부에 위치한 중구와 동구의 동들에서 높은 값이 나타나고 있다.

SSSM에서 두드러지는 공간적 체제는 광범위한 지역에서 나타나고 있는 분리-집중 경향과 동북지역에 넓게 분포하는 분리-혼재 경향이다. SSAM에 의해 유의한 것으로 판명된 동은 모두 10개인데, 7개는 분리-집중에, 3개는 분리-혼재에 속하며 중첩-혼재와 중첩-집중은 발견되지 않았다. 대전은 서울, 대구와 마찬가지로 고학력 집중지의 공간 클러스터가 서구의 북구에서 유성구로 이어지는 지역에 선명하게 발견된다. 서구의 분리-집중 클러스터는 둔산2동, 둔산1동, 합평3동, 삼천동이 연결된 지역이고, 유성구의 분리-집중 클러스터는 온천2동과 신천동이 연결되어 형성된 것이다. 분리-혼재의 전형적인 특성을 보이는 동은 대덕구 대화동인데, 주변 동 중 많은 동이 분리의 경향을 보이는 하지만 그 편차가 심하고(예를 들어 유성구 전민동과 신성동), 대덕구 중리동처럼 거주지 중첩 경향이 매우 강한 동과도 연결해 있다.

7) 울산

울산의 경우(그림 7 참조), 고학력 집단 집중지가 남구에서 주로 나타나고 있음을 알 수 있다. 2.0 이상의 값을 보이는 동은 남구의 삼산동(4.12)과 옥동(3.98)인데, 나머지 동과는 월등한 차이를 보여주고 있다. 남구에는 삼산동과 옥동 외에 무거2동과 신정2동도 1.0 이상의 높은 집중도를 나타내고 있다. 저학력 집단에서 높은 값을 보이는 지역이 남구, 중구, 동구, 북구에 고립 분산적으로 분포하고 있다.

SSSM을 살펴보면, 울산은 다른 도시와는 달리 분리-집중이 지역의 압도적인 비중을 차지하지 않는 특징을 볼 수 있다. 중심부에서 서쪽 방향으로 분리-혼재가 나타나며 외곽지역에 분리-혼재와 분리-집중이 나타나고 있다. SSAM을 보면, 모두 7개의 동이 유의성이 있는 것으로 확인되었는데, 분리-혼재가 3개, 분리-집중이 2개, 중첩-집중이 2개로 나타났다. 분리-집중은 남구 옥동과 중구 태화동에서 고립분산적으로 나타나고 있다. 남구 삼산동은 분리-혼재의 전형적인

특성을 나타내는데, 자체적으로는 매우 높은 고학력 집단 집중지이지만 주변의 동들은 미미한 수준의 분리 경향을 보이는 동이거나 남구 야음2동처럼 중첩의 경향이 나타나는 동도 존재한다.

## 5. 결론 및 토론

이 논문의 주된 목적은 국지적 공간 분리성 측도를 이용한 탐색적 공간데이터 분석을 통해 거주지 분화 연구에서 공간통계학적 접근이 가지는 의의를 논증하는 것이었다. 중요한 결과는 다음과 같다. 첫째, 이미 개발된 전역적 공간 분리성 측도로부터 국지적 공간 분리성 측도를 도출하였다. 둘째, 두 가지 유의성 검정을 위한 가정, 즉 총체적 랜덤화 가정과 조건적 랜덤화 가정에 기반한 가설검정 방법을 제시하였다. 셋째, 측도와 유의성 검정을 바탕으로 한 탐색적 공간데이터 분석 기법으로서 ‘공간 분리성 산포도 지도(SSSM)’와 ‘공간 분리성 이레지 지도(SSAM)’를 제시하였다. 부가적으로 각 인구 집단 별 집중도에 대한 표준화 지표도 제시되었다. 넷째, 이 기법을 우리나라 7대 도시의 고학력 집단과 저학력 집단간 거주지 분화에 적용한 결과, 특히 공간적 클러스터와 공간적 특이점을 확인하는데 유용성이 있는 것으로 판단되었다.

그러나 이 연구는 몇 가지 중요한 한계점도 드러내었다. 첫째, 전역적 통계량에 비해 국지적 통계량은 그것의 해석에 있어 많은 모호함을 가지고 있는 것으로 판명되었다. 이러한 해석상의 문제점은 이 통계량이 이변량이라는 점과 Geary의 수식을 바탕으로 했다는 것에서 파생된 것으로 보인다. 즉, 앞에서 언급한 것처럼, 이변량 측도는 네 가지 연관을 동시에 고려하게 되는데, 그 네 가지 연관들의 복잡한 관련성이 단일한 수치로 표현되기 때문에 도출 과정과 도출된 측도에 대한 직관적이 해석이 어려운 것이다. Geary의 수식은 기본적으로 중심 공간단위와 주변 공간단위간의 차이에 기반하고 있는데, 이 점 또한 해석의 어려움을 가중시킨다. 즉, 각 공간단위가 지니고 있는 원 자료의 수치에는 상관없이 오로지 그 수치와 주변 공간단위가

지니고 있는 수치와의 차이에만 집중하기 때문에, 중심 공간단위에서의 거주지 분리/중첩의 상황이 매우 상이하더라도 주변 공간단위와의 수치적 차이가 유사하다면 거의 비슷한 국지 통계량이 도출된다. 둘째, 위의 문제와 관련된 것으로서, 본 연구에서 제시된 국지적 통계량은 클러스터의 확인에는 높은 수행성을 보여 주지만, 확인된 클러스터가 어느 집단이 지배적인 클러스터인가에 대한 정보는 제공하지 않는다. 사례 연구에서 나타난 것처럼, 그것을 알기 위해서는 각 집단 별 집중도 지도와 같은 부가 정보의 도움을 받아야만 한다.

이러한 문제점을 해소하는데 local Lee's  $L_i$ 가 하나의 대안으로 제시될 수 있다(Lee, 2001a; 2001b; 2008). 이 통계량은 중심 공간단위와 주변 공간단위 간을 분리하지 않고 두 개를 결합하여 일종의 국지적 세트를 상정한다. 이렇게 함으로써 네 종류의 연관을 두 종류의 연관, 즉 두 개의 국지적 세트 간의 관계로 치환하는 것이다. 이를 바탕으로 관련 탐색적 공간데이터 분석 기법을 적용시키면 보다 간명한 형태의 분석이 가능할 것으로 기대된다(적용 사례를 위해 Lee(2004a, 2004b)를 참고할 수 있다). ESDA를 실질적인 연구에 적용하는 것이 한계를 가지는 것은 이러한 분석 도구 자체의 한계뿐만 아니라 데이터가 주어지는 공간단위, 즉 공표구의 크기와의 관련되어 있다. 울산을 비롯해 광주, 대전의 경우에서조차 공간단위의 수가 적고 면적이 큰 공간단위가 다수 존재하는데 이러한 경우 패턴 확인을 주 목적으로 개발된 ESDA 기법은 큰 효용성을 발휘하기 어렵다. 이것은 공간데이터 분석이 공간단위의 수, 크기, 형태 등을 포괄하는 연구 세팅에 의존적임을 여실히 보여주는 것이다(Boots and Tiefelsdorf, 2000).

두 편으로 나뉘어 제시된 이 연구는 공간데이터 분석이라는 방법론적 진보가 구체적인 연구 과제의 수행을 어떻게 향상시킬 수 있는지를 논증하기 위해 기획된 것이다. 공간통계학적 관점에서 아직 해결해야 할 많은 난제가 존재하고, 소지역 통계의 이용가능성이라는 측면에서 공공적으로 향상되어야 할 국내 데이터 하부구조의 문제도 여전히 존재하고, 연구 수행을 위한 소프트웨어 환경이라는 측면에서도 많은 과제가 존

재한다. 그럼에도 불구하고, ESDA 방법론을 개발하고, 그것을 실질적인 지리적 질문에 대한 해답을 구하기 위해 적용하는 것은 지리학적인 공간통계 분석, 혹은 더 나아가 진정으로 지리학적인 지리정보과학(GIScience)을 정초하는데 중요한 역할을 담당할 것이다.

## 文獻

- 이상일, 2007, “거주지 분화에 대한 공간통계학적 접근 (I): 공간 분리성 측도의 개발,” *대한지리학회지*, 42(4), 616-631.
- 최은영, 2004, 서울시 거주지 분리 심화와 교육환경의 차별화, 서울대학교 대학원 사회교육과(지리전공) 박사학위논문.
- Anselin, L., 1995, Local indicators of spatial association: LISA, *Geographical Analysis*, 27(2), 93-115.
- Anselin, L., 1996, The Moran scatterplot as an ESDA tools to assess local instability in spatial association, in Fischer, M., Scholten, H., and Unwin, D. (eds.), *Spatial Analytical Perspectives on GIS*, Taylor & Francis, London, 111-125.
- Anselin, L., 1998, Exploratory spatial data analysis in a geocomputational environment, in Longley, P. A., Brooks, S. M., McDonnell, R., and MacMillan, B. (eds.), *Geocomputation: A Primer*, John Wiley & Sons, Chichester, West Sussex, 77-94.
- Anselin, L., Syabri, I., and Smirnov, O., 2002, Visualizing multivariate spatial correlation with dynamically linked windows, in Anselin, L. and Rey, S. (eds.), *New Tools for Spatial Data Analysis: Proceedings of the Specialist Meeting*, Center for Spatially Integrated Social Science (CSISS), University of California, Santa Barbara.
- Bailey, T. C., 1994, A review of statistical spatial analysis in geographical information systems, in Fotheringham, A. S. and Rogerson, P. (eds.), *Spatial Analysis and GIS*, Taylor & Francis, London, 13-44.
- Boots, B. and Tiefelsdorf, M., 2000, Global and local spatial autocorrelation in bounded regular tessellations, *Journal of Geographical Systems*, 2(3), 319-348.
- Brown, L. A. and Chung, S.-Y., 2006, Spatial segregation, segregation indices and the geographical perspective, *Population, Space and Place*, 12(2), 125-143.
- Cliff, A. D. and Ord, J. K., 1981, *Spatial Processes: Models & Applications*, Pion Limited, London.
- Chung, S.-Y. and Brown, L. A., 2007, Racial/ethnic residential sorting in spatial context: Testing the exploratory frameworks, *Urban Geography*, 28(4), 312-339.
- Feitosa, F. F., Camara, G., Monteiro, A. M. V., Koschitzki, T., and Silva, M. P. S., 2007, Global and local spatial indices of urban segregation, *International Journal of Geographical Information Science*, 21(3), 299-323.
- Fischer, M. M. and Nijkamp, P., 1992, Geographic information systems and spatial analysis, *Annals of Regional Science*, 26(1), 3-17.
- Fotheringham, A. S., 1997, Trends in quantitative methods I: stressing the local, *Progress in Human Geography*, 21(1), 88-96.
- Fotheringham, A. S., 2000, Context-dependent spatial analysis: a role for GIS?, *Journal of Geographical Systems*, 2(1), 71-76.
- Fotheringham, A. S. and Charlton, M., 1994, GIS and exploratory spatial data analysis, an overview of some research issues, *Geographical Systems*, 1(4), 315-327.
- Getis, A. and Ord, J. K., 1996, Local spatial statistics: an overview, in Longley, P. and Batty, M. (eds.), *Spatial Analysis: Modelling in a GIS Environment*, GeoInformation International, Cambridge, 261-277.
- Good, I. J., 1983, The philosophy of exploratory data analysis, *Philosophy of Science*, 50(2), 283-295.
- Goodchild, M. F., Haining, R., Wise, S., et al., 1992, Integrating GIS and spatial data analysis: problems and possibilities, *International Journal of Geographical Information Systems*,

- 6(5), 407-423.
- Hubert, L. J., 1984, Statistical applications of linear assignment, *Psychometrika*, 49(4), 449-473.
- Hubert, L. J., 1987, *Assignment Methods in Combinatorial Data Analysis*, Marcel Dekker, New York.
- Lee, S.-I., 2001a, Developing a bivariate spatial association measure: An integration of Pearson's  $r$  and Moran's  $I$ , *Journal of Geographical Systems*, 3(4), 369-385.
- Lee, S.-I., 2001b, *Spatial Association Measures for an ESDA-GIS Framework: Developments, Significance Tests, and Applications to Spatio-temporal Income Dynamics of U.S. Labor Market Areas, 1969-1999*, Ph.D. Dissertation, Department of Geography, The Ohio State University.
- Lee, S.-I., 2004a, Exploratory spatial data analysis of  $\sigma$ -convergence in the U.S. regional income distribution, 1969-1999, *Journal of the Korean Urban Geographical Society*, 7(1), 79-95.
- Lee, S.-I., 2004b, Spatial data analysis for the U.S. regional income convergence, 1969-1999: A critical appraisal of  $\beta$ -convergence, *Journal of the Korean Geographical Society*, 39(2), 212-228.
- Lee, S.-I., 2004c, A generalized significance testing method for global measures of spatial association: An extension of the Mantel test, *Environment and Planning A*, 36(9), 1687-1703.
- Lee, S.-I., 2005, Between the quantitative and GIS revolutions: Towards an SDA-centered GIScience, *Journal of Geography Education*, 49, 268-284.
- Lee, S.-I., 2008, A generalized randomization approach to local measures of spatial association, *Geographical Analysis*, under revision.
- Leung, Y., Mei, C.-L., and Zhang, W.-X., 2003, Statistical test for local patterns of spatial association, *Environment and Planning A*, 35(4), 725-744.
- Maly, M. T., 2000, The neighborhood diversity index: a complementary measure of racial residential settlement, *Journal of Urban Affairs*, 22(1), 37-47.
- Oden, 1995, Adjusting Moran's  $I$  for population density, *Statistics in Medicine*, 14(1), 17-26.
- Openshaw, S., 1990, Spatial analysis and geographical information systems: a review of progress and possibilities, in Scholten, H. and Stillwell, J. (eds.), *Geographical Information Systems for Urban and Regional Planning*, Kluwer, Dordrecht, 153-163.
- Openshaw, S. and Clarke, G., 1996, Developing spatial analysis functions relevant to GIS environments, in Fischer, M., Scholten, H., and Unwin, D. (eds.), *Spatial Analytical Perspectives on GIS*, Taylor & Francis, London, 21-37.
- Rogerson, P. A., 1999, The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic, *Geographical Analysis*, 31(1), 130-147.
- Sokal, R. R., Oden, N. L., and Thomson, B. A., 1998, Local spatial autocorrelation in a biological model, *Geographical Analysis*, 30(4), 331-354.
- Tiefelsdorf, M., 1998, Some practical applications of Moran's  $I$ 's exact conditional distribution, *Papers in Regional Science*, 77(2), 101-129.
- Tukey, J. W., 1977, *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.
- Unwin, D. J., 1996, GIS, spatial analysis and spatial statistics, *Progress in Human Geography*, 20(4), 540-551.
- Wise, S., Haining, R., and Signoretta, P., 1999, Scientific visualization and the exploratory analysis of area data, *Environment and Planning A*, 31(10), 1825-1838.
- Wong, D. W. S., 2002, Modeling local segregation: a spatial interaction approach, *Geographical & Environmental Modelling*, 6(1), 81-97.
- Wong, D. W. S., 2003, Spatial decomposition of segregation indices: A framework toward measuring segregation at multiple levels, *Geographical Analysis*, 35(3), 179-194.
- 교신: 이상일, 151-748, 서울특별시 관악구 신림동 산 56-1, 서울대학교 사범대학 지리교육과(이메일:

si\_lee@snu.ac.kr, 전화: 02-880-9028, 팩스: 02-882-9873)

Correspondence: Sang-Il Lee, Department of Geography Education, College of Education, Seoul National University, San 56-1, Silim-dong, Gwanak-gu,

Seoul 151-748, Korea (e-mail: si\_lee@snu.ac.kr, phone: +82-2-880-9028, fax: +82-2-882-9873)

최초투고일 08. 03. 18.

최종접수일 08. 03. 25.