대한지리학회지 제42권 제4호 2007(616~631)

거주지 분화에 대한 공간통계학적 접근 (I): 공간 분리성 측도의 개발*

이상일**

A Spatial Statistical Approach to Residential Differentiation (I): Developing a Spatial Separation Measure*

Sang-Il Lee**

요약 : 거주지 분화 현상은 도시적 삶의 공간성을 파악하는데 본질적인 요소이기 때문에 도시학 연구에서 오랫동안 주목을 받아왔다. 거주지 분화 현상에 대한 연구 과제 중의 하나가 상이한 두 집단이 얼마나 공간적으로 분리되어 있는지를 측정하는 문제이다. 이러한 측면에서 가장 널리 사용되어온 것이 상이지수(index of dissimilarity)인데, 이 지수는 거주지 분리의 '불균등성(unevenness)' 은 측정할 수 있지만, 공간적 '집중도(clustering)' 는 측정하지 못하는 단점을 갖고 있다. 이러한 단점을 극복하기 위해 제안되어 온 '공간적 격리 지수(spatial indices of segregation)' 역시 가설검정 절차를 제시하지 못하고 최근의 공간통계학 연구 성과를 수용하지 못하는 등의 단점을 가지고 있다. 이러한 의미에서 본 논문의 주된 연구 목적은 새로운 '공간 분리성 측도(spatial separation measure)' 를 개발하는 것이다. 이 공간 분리성 측도는 상이한 인구 집단이 거주 공간에 얼마나 불균등하게 분포하고 있는지에 대한 것뿐만 아니라 그러한 불균등 분포가 보여주는 공간적 의존성의 정도 까지도 측정하는 새로운 통계량이다. 주요 연구 결과는 다음과 같다. 첫째, 기존의 '공간 연관성 측도(spatial association measures)' 와 '공간적 카이-스퀘어 통계량(spatial chi-square statistics)' 을 통합하여 새로운 측도를 개발했으며, 일반화된 랜덤화 검정법을 적용해 측도에 대한 유의성 검정법을 제시하였다. 둘째, 개발된 측도와 유의성 검정법을 우리나라 7대 도시의 학력 집단 간 거주지 분리 현상에 적용함으로써, 연구 방법론으로서의 유용성을 확인하였다.

주요어 : 공간 분리성 측도, 거주지 분리, 상이지수, 공간통계학, 공간적 의존성, 공간 연관성 측도, 공간적 카이-스퀘어 통계량

Abstract : Residential differentiation is an academic theme which has been given enormous attention in urban studies. This is due to the fact that residential segregation can be seen as one of the best indicators for socio-spatial dialectics occurring on urban space. Measuring how one population group is differentiated from the other group in terms of residential space has been a focal point in the residential segregation studies. The index of dissimilarity has been the most extensively used one. Despite its popularity, however, it has been accused of inability to capture the degree of spatial clustering that unevenly distributed population groups usually display. Further, the spatial indices of segregation which have been introduced to edify the problems of the index of dissimilarity also have some drawbacks: significance testing methods have never been provided; recent advances in spatial statistics have not been extensively exploited. Thus, the main purpose of the research is to devise a spatial separation measure which is expected to gauge not only how unevenly two population groups are distributed over urban space, but also how much the uneven

* 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

distributions are spatially clustered (spatial dependence). The main results are as follows. First, a new measure is developed by integrating spatial association measures and spatial chi-square statistics. A significance testing method based on the generalized randomization test is also provided. Second, a case study of residential differentiation among groups by educational attainment in major Korean metropolitan cities clearly shows the applicability of the analytical framework presented in the paper.

Key Words : spatial separation measure, residential segregation, index of dissimilarity, spatial statistics, spatial dependence, spatial association measures, spatial chi-square statistics

1. 문제 제기 및 연구 목적

도시 내 거주 집단 간 공간적 배제의 양상은 도시사회지리학의 전통적인 주제일 뿐만 아니라 인접 사회과학의 다양한 분야, 특히 도시사회학 혹은 도시경제학의 주요한 주제 중의 하나이기도 하다. 상이한 사회경제적 특성을 갖는 인구 집단은 도시공간 내에서 상이한 지점을 점유하는 경향이 있으며, 그러한 집단 간 공간적 분리는 다시 사회경제적 분리를 재생산하는데 기여한다. 특정한 집단들 사이에서 거주지 분리의 정도가 높다는 것은 두 집단이 다른 집단들에 비해 집단 간 배제의 정도가 높다는 것을 의미하며, 특정 도시의 집단 간 거주지 분리의 정도가 다른 도시에 비해 높다는 것은 그 도시의 거주 공간구조가 상대적으로 더 분절화되어 있다는 것을 의미한다. 따라서 거주지 분화의 양상은 도시 공간 상에서 발생하는 사회-공간적 역동성에 있어서 핵심적인 부분을 차지한다.

이와 같이 사회-공간적 프로세스를 이해하는 데 있어 거주지 분화가 궁극적인 것이라면, 집단 간 거주지 분리의 정도에 대해 신뢰할 만한 측도(measure)를 제공하는 것 역시 매우 중요하다(Morrill, 1991). 왜냐하면 집단 간 거주지 분리의 정도를 계량적으로 측정하는 것은 거주지 분화 현상에 대한 기술적 요약을 넘어 거주지 분화의 공간성(spatiality)에 대한 통찰력을 제시해 줄 수 있기 때문이다. 다시 말해서, 거주지 분리는 명백히 공간적 현상이기 때문에, 그것을 측정하는 도구 역시 거주지 분리의 공간적 차원을 분명히 드러내는 형태로 정식화되어야 한다는 것이다. 이러한 의미에서, 본 연구는 개념적으로나 통계학적으로 신뢰할 만한 측도를 개발하기 위해서는 최근 급속히 발전하고 있는 공간통계학(spatial statistics)의 연구성과에 기반하여 거주지 분리에 대한 기존의 지수들을 재조명할 필요가 있다는 데 기본적인 문제의식을 갖는다.

이러한 측면에서, 거주지 분리 정도를 측정하기 위해 흔히 사용되는 상이지수(相異指數, index of dissimilarity)는 몇 가지 문제점을 갖고 있다. 상이지수는 명백히 공간적인 현상인 거주지 분리에 대한 비공간적인 지수이다. 즉, 전혀 다른 수준의 공간적 분리에 대해 동일한 상이지수가 도출될 수 있다(White, 1983; Wong, 1993). 또한 상이지수는 단순히 수학적 지수일 뿐 통계적 측도가 아니다. 즉, 지수에 대한 신뢰할 만한 유의성 검정법이 제시되어 있지 않다. 공간적 격리 지수(spatial indices of segregation)의 개발을 위한 최근의 노력들은 위의 상이지수의 한계를 극복하려는 시도로 높이 평가되어야 한다. 그러나 제안된 공간적 격리 지수의 경우도 몇 가지 문제점을 갖고 있다. 우선, 다양한 공간적 격리 지수가 제시되었지만 합의될만한 수준의 정교한 지수는 제시되지 못했다. 또한 공간적 격리 지수들 역시 여전히 지수일 뿐 통계적 측도가 아니다.

위에 제시된 문제의식을 바탕으로, 본 연구의 목적은 공간통계학의 최근 발전을 거주지 분화 연구에 적용하여, 새로운 측도를 개발하고 유의성 검정법을 제시하여, 이를 우리나라 7대 대도시의 거주지 분화에 대한 분석적 연구에 적용하는 것이다. 보다 구체적인 연구과제는 다음과 같다.

첫째, 새로운 공간 분리성 측도(spatial separation measure)와 유의성 검정법을 개발한다. 이를 위해, 기존의 공간적 격리 지수를 비판적으로 검토하고, 공간통계학에서 개발된 공간 연관성 측도(spatial association measures)와 공간적 카이-스퀘어 통계량

(spatial chi-square statistics)을 참조한다. 또한 유의성 검정을 위해 '일반화된 랜덤화 접근법(generalized randomization approach)' (Lee, 2004)의 적용가능성을 검토한다. 이러한 논의를 바탕으로 공간 분리성 측도와 그것의 통계적 속성, 그리고 가설검정 방법에 대한 핵심적 사항이 개발될 것이다.

둘째, 위의 방법론을 우리나라 7대 도시(서울, 부산, 인천, 대구, 광주, 대전, 울산)의 동별 거주지 분리 양상에 적용한다. 이 사례연구에서는 40~59세 인구를 대상으로 학력 집단 간 거주지 분리가 연구된다. 거주지 분리의 기준으로 학력자본을 대상으로 삼은 것은 거주지 분리가 단순히 경제적 자본에 의해서만 규정되는 것이 아니라 사회문화적 자본에 의해서도 영향 받는다는 점을 고려한 결과이며, 40~59세의 인구만을 선정한 것은 40세를 학력 자본이 거주지 선택을 통해 실현되는 시점으로 보았기 때문이고, 60세 이상을 배제한 것은 연령 요소의 작동을 회피하기 위한 조치이다. 인구를 고학력, 중학력, 저학력으로 구분하고 세 집단 간 거주지 분리 정도를 측정할 것이다.

2. 기존 거주지 분리 지수에 대한 비판적 고찰

1) 상이지수(index of dissimilarity)의 문제

상이지수는 두 집단 사이에서 완전한 거주 분포의 균형을 맞추기 위해 얼마만큼의 사람이 이동해야 하는 지를 측정함으로써 현 거주 분포 상에서의 두 집단 간 거주 불일치성을 수치화한 것이다(Duncan and Duncan, 1955). 상이지수를 구하는 수식은 아래와 같다.

$$D = \frac{1}{2}\sum_i \left| \frac{X_i}{X} - \frac{Y_i}{Y} \right| \qquad (1)$$

여기에서 X는 전체 지역 X집단의 인구 수, X_i는 i지역에 거주하는 집단 인구수를 나타내므로, X_i/X는 i지역 X집단 인구의 전체 X집단 인구에 대한 비율을 의미한다.

이러한 상이지수가 갖는 한계에 대해서는 사회학자 및 인구학자들에 의해 지적되었지만, 지리학적인 혹은 공간분석적인 측면에서 보다 중요한 것은 상이지수가 가지고 있는 몰공간성을 지적하는 것이다. 즉, 상이지수는 거주지 분리의 '공간성'을 측정하지 못한다. 이를 명시적으로 보여주는 것이 White(1983: 1010)가 '체커보드 문제(checkerboard problem)' 라고 부른 것이다. 이것은 공간적 분포 패턴이 가지고 있는 공간적 의존성(spatial dependence) 혹은 공간적 자기상관(spatial autocorrelation)의 문제를 지적하는 것이다. 그림 1은 100개의 공간단위로 구성된 가상적 연구지역에 드러난 두 개의 상이한 거주지 분리 양상을 보여주고 있다. 음영으로 처리된 부분은 A집단이 100% 거주하는 지역이고, 음영으로 처리되지 않은 부분은 B집단이 100% 거주하는 지역이라고 할 때, 오른쪽 패턴이 왼쪽 패턴에 비해 격리의 '공간적' 양상이 훨씬 더 두드러짐에도 불구하고 동일한 상이지수 값이 산출된다. 즉, 전자는 극단적인 음(-)적인 공간적 의존성을 보이지만 후자는 현저한 양(+)적인 공간적 의존성을 보여주고 있는데, 상이지수는 이를 구분하지 못하는 것이다. Massey and Denton(1988)의 용어를 빌면, 상이지수는 거주지 분리의 '불균등성(unevenness)' 은 측정할 수 있지만, 공간적 '집중도(clustering)' 는 측정하지 못한다. 즉, 두 공간적 패턴은 불균등성에서는 같지만, 집중도에서는 현저히 다른 것이다.

이러한 '집중도' 의 차이가 가지는 의미는 명백하다. 거주지 분리를 사회경제적 배제의 중요한 수단이자 결과라고 할 때, 사회경제적 배제의 심각성은 오른쪽 패턴에서 훨씬 더 현저하다. 또한 집단간 이질성의 심화와 갈등의 증폭 역시 오른쪽 패턴에서 훨씬 더 격렬할 것이 자명하다. 집단 간 사회경제적 분리의 공간구조가 사회경제적 과정 그 자체를 매개한다는 점을 전제할 때 이 두 경우는 거주지 분화의 본질적인 성격 자체가 완전히 다른 것이다. 따라서 상이지수는 거주지 분리의 공간적 측면을 제대로 반영하고 있지 못하며, 그림 1에서 제시된 두 패턴을 명백히 구분해 줄 수 있는 공간적 측도를 개발하는 것은 매우 시급한 과제가 된다.

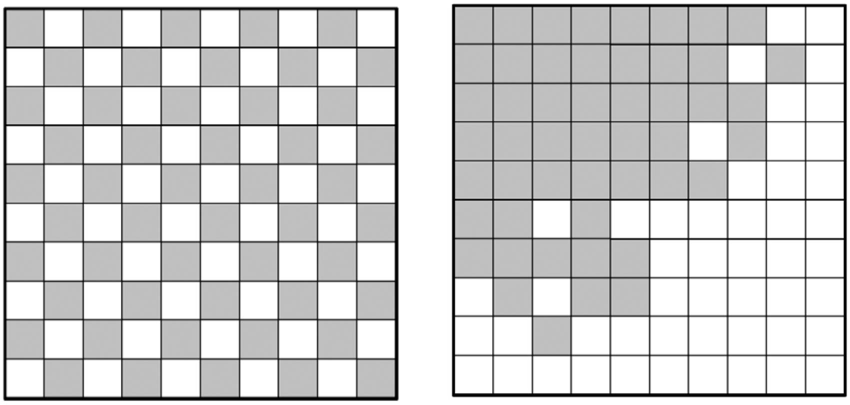


그림 1. White의 체커보드 문제(최은영, 2004, 18)

2) 공간적 격리 지수(spatial indices of segregation)의 등장과 한계

지리학자를 중심으로 위에서 제시된 상이지수의 몰공간성에 대한 지적이 있어왔으며, 그에 따라 공간화된 격리 지수가 제안되었다. 그 중 중요한 것으로, Jakubs(1981)의 거리-기반 격리 지수(DBI: distance-based segregation index), Morgan(1983)의 수정된 거리-기반 지수(MDBI: modified distance-based index), White(1983; 1986)의 P 지수, Massey and Denton(1988)의 ACL, Morrill(1991)의 $D(adj)$, Wong(1993)의 $D(w)$와 $D(s)$, Waldorf(1993)의 할당-기반 격리 지수(relocation-based segregation index), Charkravorty(1996)의 근린주구 불균등 지수(NDI; neighbourhood disparity index), Lee and Culhane (1998)의 집중 지수(clustering index), Wu and Sui(2001)의 lacunarity-기반 격리 지수, Dawkins (2004)의 공간적 지니 지수, Feitosa *et al.*(2007)의 일반화된 공간적 상이 지수(generalized spatial dissimilarity index) 등이 있다.

여기에서는 상이지수를 변형하는 형태를 취하고 있는 Morrill과 Wong의 지수에 집중하고자 한다. 각 지수는 다음과 같이 주어진다.

* Morrill(1991) : $D(adj) = D - \dfrac{\sum_i\sum_j |c_{ij}(z_i - z_j)|}{\sum_i\sum_j c_{ij}}$

* Wong(1993) : $D(w) = D - \frac{1}{2}\sum_i\sum_j w_{ij}\,|z_i - z_j|$

* Wong(1993) :

$$D(s) = D - \frac{1}{2}\sum_i\sum_j w_{ij}\,|z_i - z_j| \times \frac{1/2\lfloor (P_i/A_i) + (P_j/A_j)\rfloor}{\mathrm{MAX}(P/A)}$$

여기서 D는 상이지수를 의미하며, z_i는 i번째 공간단위 내에서의 특정 집단의 비중을, z_j는 i와 공간적으로 인접한 공간단위에서의 특정 집단의 비중을 의미한다. P와 A는 공간단위의 형상을 나타내는 지표들이다. 이 공간적 지수들은 모두 상이지수에 기반을 두고서, 특정 집단의 비율분포의 공간적 의존성 혹은 집중도를 고려하고 있다는 공통점을 갖고 있다. 보다 구체적으로 표현하면, 공간적 격리 지수는 상이지수에서 공간적 의존성 부분을 차감하는 형태를 취한다. 만일 특정한 집단의 분포 패턴이 양(+)적인 공간적 의존성 혹은 공간적 자기상관을 보인다면, 공간적 의존성을 측정하는 수식 부분은 낮은 값을 가질 것이므로, 결국 상이지수는 적게 삭감될 것이다. 반대로, 음(-)적인 공간적 의존성 혹은 공간적 자기상관을 보인다면, 공간적 의존성 부분의 값이 커져 전체적으로 상이지수의 값이 낮아지게 된다. 결국 변형된 D값이 높을수록 공간적 격리의 수준이 높다고 말할 수 있다.

공간적 격리 지수가 어떻게 작동하는지 보여주기 위해 상이지수가 1로 동일하지만 공간적 의존성은 다른 세 가지 패턴에 대해 위에 제시된 지수값을 적용한 것을 정리하면 그림 2와 같다(Wong, 1993). 음영으로

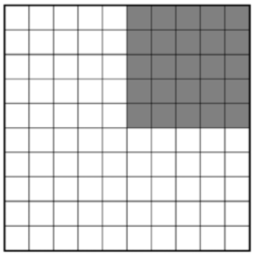
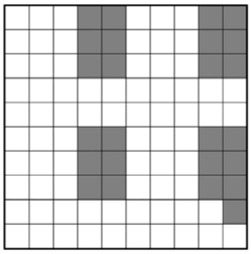
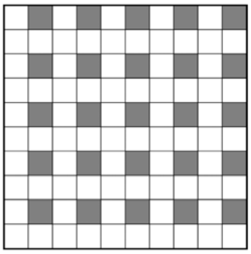
공간적 패턴	*D*	*D(adj)*	*D(w)*	*D(s)*
	1	0.94	0.94	0.95
	1	0.83	0.83	0.84
	1	0.50	0.50	0.54

그림 2. 공간적 격리 지수(Wong, 1993, 563, Figure 2를 수정함)

처리된 부분은 대상 집단이 100% 거주하는 구역이다. 지수값에서 보듯이 제안된 공간적 지수들은 세 패턴에 대해 모두 다른 값을 산출하고 있으며, 공간성 의존성이 높은 패턴에 높은 값이 주어져 있음을 알 수 있다.

이러한 공간적 격리 지수는 그 가능성에도 불구하고 다음의 세 가지 점에서 한계가 있다. 첫째, 여전히 비공간적인 상이지수에 의존적이며, 제안된 지수 중 어느 것이 보다 효율적인지 밝혀져 있지 않다. 위의 세 지수는 엄밀히 말해 공간 근접성 행렬(spatial proximity matrix)을 어떻게 구성하느냐에 따라 구분될 뿐 본질적으로는 아무런 차이가 없다. 둘째, 공간적 격리 지수는 여전히 지수일뿐 측도나 통계량으로서의 지위에 있지 않다. 지수가 통계량이기 위해서는 그 지수의 표본분포(sampling distribution)가 밝혀지거나, 최소한 기대값과 분산을 계산할 수 있는 수식이 제시되어야 한다. 실질적으로 상이지수 자체에 대한 유의성 검정도 최근에 와서야 제시되고 있는 실정이다(Ransom, 2000). 셋째, 국지적인 공간적 격리 지수를 개발하려는 시도가 있기도 했지만(Maly, 2000; Wong, 2002; 2003; Feitosa *et al.*, 2007), 이러한 경우도 공간 연관성 측도(spatial association measures)의 연구에서 제시되고 있는 전역적 측도와 국지적 측도 간에 존재하는 가법성 요구(additivity requirement), 즉, 국지적 측도의 평균은 전역적 측도와 같다는 조건을 만족시키지 못한다. 이것은 국지적 패턴을 전역적 경향과 관련 지어 설명할 수 없다는 사실을 의미한다. 따라서 새로이 개발될 측도는 전역적 측도로부터 국지적 측도가 직접적으로 연역될 수 있는 형태를 취해야

만 한다. 새로운 측도를 기존의 공간적 격리 지수와 구분하고 '격리' 보다는 보다 중립적인 의미를 함축하기 위해 새로이 개발될 지수에 대해 공간 분리성 측도라는 이름을 부여하기로 한다.

3. 공간 분리성 측도(spatial separation measure)의 개발

1) 공간 연관성 측도와 공간적 카이-스퀘어 통계량에 대한 검토

(1) 공간 연관성 측도(spatial_association measures)

앞에서 언급한 것처럼, 새롭게 개발될 공간 분리성 측도는 특정 집단의 공간적 분포가 보여주는 공간적 의존성 혹은 공간적 자기상관에 대한 고려를 포함해야만 한다. 이는 공간 분리성 측도의 개발이 공간 연관성 측도와 필연적으로 결합되어 있음을 의미한다. 거주지 분리 연구에서 일변량(univariate) 공간 연관성 측도를 사용하는 것의 중요성이 지적되긴 했지만(Massey and Denton,1988; Wong, 1997; Lee, 2001; Dawkins, 2004; 최은영, 2004; Brown and Chung, 2006), 정교한 연구의 단계로 진전된 적은 없었다. 또한 주로 Moran 통계량에 집중하는 경향이 있어, 연구 대상에 따라 상이한 공간 연관성 측도가 사용될 필요성이 강조되지 못하고 있는 실정이다. 특히, Moran 통계량이 공간적 의존성을 측정하기 위해 인접 값들 간의 공분산(covariance)를 사용하는데 반해, Geary 통계량은 값들 간의 '차이'를 사용하고 있다는 점에 주목할 필요가 있다. 이것은 상이지수나 공간적 격리 지수와 유사성을 갖는 것으로 오히려 Geary의 통계량이 거주지 분리 연구에서는 더 적절할 수 있음을 시사하는 것이다.

그러나 새로운 공간 분리성 측도는 이변량(bivariate) 공간 연관성 측도에서 더 많은 통찰력을 얻어야만 한다. 왜냐하면, 두 집단간 거주지 분리란 두 집단의 공간적 분포 패턴 간의 비교를 의미하는 것이기 때문이고, 이변량 공간 연관성 측도가 그러한 公形(co-patterning)의 정도 혹은 이변량 공간적 의존성의 정도를 측정하는 지표로 제시되고 있기 때문이다(Lee, 2001). 현재 CM(Cross-Moran)(Wanternberg, 1985)과 Lee's L(Lee, 2001)이 제시되어 있는데, CM은 그 자체로 많은 문제점을 가지고 있기 때문에, Lee's L이 보다 합리적인 거주지 분리 측정의 대안으로 제시될 수 있을 것으로 판단된다.

$$L = \frac{n}{\sum_i(\sum_j v_{ij})^2} \cdot \frac{\sum_i[(\sum_j v_{ij}(x-\bar{x})) \cdot (\sum_j v_{ij}(y-\bar{y}))]}{\sqrt{\sum_i(x_i-\bar{x})^2}\sqrt{\sum_i(y_i-\bar{y})^2}} \quad (2)$$

그러나 새로이 개발될 공간 분리성 측도는 Geary의 일변량 통계량을 이변량으로 확장한 이변량 Geary를 개발함으로써 보다 합리적인 토대를 마련할 수 있을 것으로 판단된다. 왜냐하면 앞에서 언급한 것처럼, 거주 분포가 보여주는 공간적 의존성은 일변량 Geary 통계량에서 살펴본 것처럼, 인접한 값들의 차이를 통해 보다 의미 있게 측정될 수 있다고 판단되기 때문이다. 공간 연관성 측도가 시사하는 또 다른 중요한 사항은 위에서 제시된 전역적 측도가 간단한 분해과정을 통해 국지적 측도로 용이하게 전환될 수 있다는 사실이다.

(2) 공간적 카이-스퀘어 통계량(spatial chi-square statistics)

새로운 공간 분리성 측도를 개발하기 위해 빈도분포에서의 일치성을 다루는 적합도 분석(goodness-of-fit) 혹은 보다 넓은 의미의 범주 데이터 분석(categorical data analysis)과 그것을 공간 역학(疫學)(spatial epidemiology)에서 확장한 연구 성과들을 살펴볼 필요가 있다(Waller and Gotway, 2004). 왜냐하면, 거주지 분리란 기본적으로 공간단위라고 하는 범주를 가로질러 인구 집단이 어떤 빈도 분포를 보이고 있느냐의 문제를 의미하며, 그 빈도 분포가 이질적일 때 우리는 두 집단간의 거주지 분리가 존재한다고 말할 수 있기 때문이다. 이러한 측면에서 기존의 카이-스퀘어 검정을 공간적으로 확장한 Tango(1995)와 Rogerson (1999)의 연구는 시사하는 바가 크다.

어떤 속성의 지역적 빈도분포가 공간적 의존성을 보이는지 그렇지 않은지를 검정할 수 있는 방법으로 제

안된 것이 Tango 지수이다(Tango, 1995).

$$T = \sum_i \sum_j a_{ij}(r_i^X - p_i)(r_j^X - p_j) \tag{3}$$

여기에서 a_{ij}는 공간단위 간 위상 관계를 정의하는 공간 근접성 행렬의 한 요소이고, r_i^X는 특정 인구 집단 X의 인구 중 i번째 공간단위에서 분포하는 인구의 비중, 즉 열-비중(column-proportions), r_j^X는 i번째 공간단위와 인접한 공간단위에 분포하는 X집단 인구의 열-비중을 의미하고, P_i는 모든 인구 집단을 포괄한 전체 인구 중 i번째 공간단위에서 분포하는 인구의 열-비중을, P_j는 i번째 공간단위와 인접한 공간단위에 분포하는 전체 인구의 열−비중을 의미한다.

Rogerson(1999)는 $w_{ij} = a_{ij}/\sqrt{p_i p_j}$ 그리고 $w_{ii} = a_{ii}/p_i$ 라고 할 경우, 위의 식이 다음과 같이 분해됨을 보이면서 새로운 공간적 카이−스퀘어 통계량을 제안하였다.

$$\begin{aligned} R &= \sum_i \sum_j w_{ij}(r_i^X - p_i)(r_j^X - p_j) \\ &= \sum_i \frac{(r_i^X - p_i)^2}{p_i} + \sum_i \sum_{j \neq i} a_{ij} \frac{(r_i^X - p_i)(r_j^X - p_j)}{\sqrt{p_i p_j}} \end{aligned} \tag{4}$$

이 때 앞 부분은 일반적인 비공간적 카이-스퀘어 통계량이고, 뒤 부분은 바로 공간적 의존성과 관련된 부분이다. 식(3)과 식(4)가 표현하고 있는 것은 각 공간단위의 특정 집단의 열-비중을 전체 인구 집단의 열-비중으로 일종의 표준화를 한 후, 그것의 분포가 보여주는 공간적 의존성을 측정하고 있다.

이러한 공간적 카이−스퀘어 통계량이 시사하는 바는 새로이 개발될 공간 분리성 측도가 두 집단의 열-비중의 관계를 통해 정식화되어야 한다는 점이다. 일반적으로 공간 연관성 측도는 한 집단이 한 공간단위 내에서 차지하는 비중, 즉 행−비중(row-proportions)에 초점을 두고 있다. 그러나 보다 더 중요한 것은 공간적 카이-스퀘어 통계량이 공간 연관성 측도와 달리 각 공간단위의 상이한 표본 개수를 고려하고 있다는 사실이다. 비록 열−비중을 사용한다 하더라도 공간 연관성 측도의 경우 두 공간단위에서 나타나는 비중 값이 같다면, 전체 측도에 대한 기여도가 동등한 것으로 간주한다. 즉, 비중 값이 총 인구 수가 매우 큰 공간단위에서 측정된 것인지 매우 적은 공간단위에서의 측정된 것인지를 구분하지 않는다. 이러한 측면에서, 각 공간단위의 특정 인구 집단의 열−비중을 전체 인구 집단에서 각 공간단위가 차지하는 비중으로 표준화하는 것은 매우 합리적인 것으로 판단된다. 이것은 Oden(1995)이 수정된 Moran 통계량을 제시한 이유와 일맥상통하는 것이다.

(3) 공간 분리성 측도의 조건

위에서 살펴 본 것을 바탕으로 새롭게 개발될 공간 분리성 측도가 가져야 하는 조건을 설정하면 다음과 같다.

첫째, 공간 분리성 측도는 한 인구집단이 한 공간단위 내에서 갖는 비중(행−비중)이 아니라 한 인구집단이 공간단위 간에서 갖는 비중(열−비중)으로 구성되어야 한다. 이것은 두 집단이 공간 상에서 분리되어 있는가 그렇지 않은가를 평가하기 위해서는 각 공간단위 내에서 두 집단이 차지하는 비중을 비교하는 것이 아니라 두 집단이 공간단위를 가로질러 어떻게 분포하고 있는가를 비교해야 하기 때문이다. 따라서 이것은 상이지수의 기본틀을 유지한다는 것을 의미한다. X, Y 두 집단의 i 공간단위에서의 열-비중을 각각 r_i^X, r_j^Y로 표현을 할 때, 식(1)에 나타나 있는 상이지수는 다음과 같이 표현될 수 있다.

$$D = \frac{1}{2}\left| r_i^X - r_i^Y \right| \tag{5}$$

다시 말하면, 공간 분리성 측도는 행−비중 간의 상관관계가 아니라 열−비중 간의 상관관계로 규정되어야 한다. 더 나아가 각 열−비중은 공간적 카이−스퀘어 통계량이 보여준 것처럼 표준화될 필요가 있다. 사실 식(5)도 다음과 같이 주어질 수 있다.

$$D = \frac{1}{2}\left| (r_i^X - p_i) - (r_i^Y - p_i) \right| \tag{6}$$

둘째, 공간단위 간의 위상관계가 수식에 포함되어야 한다. 공간 분리성 측도는 각 공간단위 내에서 발생하는 집단 간 분리뿐만 아니라 인접 공간단위에서 발생하는 분리도 동시에 고려해야 하기 때문이다. 보다 엄밀하게 말하면 공간 분리성 측도는 네 종류의 연관성

을 동시에 고려해야 한다. 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, 이 네 가지가 동시에 고려되어야 하는 것이다. 사실 상 이 네 가지 요소를 독립적으로 동시에 고려한다는 것은 불가능하지만, 개념적인 수준에서라도 수식 내에서 함께 다루어져야만 한다. 위에서 소개된 공간적 격리 지수들도 이러한 측면을 고려하고 있기는 하지만 그 통합의 방식이 상이지수에 대한 차감효과를 측정하기 위해 소극적으로 쓰이고 있다는 데 그 한계가 있다. 즉, 비공간적 상이지수와 분리된 채 다루어질 것이 아니라 두 가지가 동시에 고려될 수 있는 통합적 수식 구조로 제시될 필요가 있는 것이다. 이러한 측면에서 이변량 공간 연관성 측도의 구조를 참조할 수 있다. 또 앞에서 언급했던 것처럼 공간 연관성 측도의 구조를 따름으로써 전역적 측도로부터 용이하게 국지적 측도를 이끌어 낼 수 있다는 부수적 이익도 취할 수 있게 된다.

2) 공간 분리성 측도의 개발

우선, 식(3)에서 주어진 Tango 지수를 이변량화하면 다음과 같이 주어질 수 있다.

$$T = \sum_i \sum_j v_{ij} [(r_i^X - p_i) - (r_j^X - p_j)] \cdot [(r_i^Y - p_i) - (r_j^Y - p_j)] \qquad (7)$$

여기서 v_{ij}는 공간 근접성 행렬(spatial proximity matrix) **V**의 요소로 두 공간단위 간 위상 관계를 규정한다. 만일 공간 근접성 행렬이 이항 연접성(binary contiguity)을 바탕으로 정의되었다면, 두 공간단위가 경계를 접하고 있으면 1의 값을, 그렇지 않으면 0의 값을 갖는다. r_i^X와 r_j^X는 위에서 규정한 정의를 따르고, r_i^Y와 r_j^Y는 또 다른 인구 집단 Y에 대한 것이다. 수식을 살펴보면, 위에서 언급한 것처럼, 우선 각 공간단위의 열-비중이 표준화되어야 함을 알 수 있다. 그리고 나서 각 공간단위에 대해 이웃 공간단위 하나를 선택한 후 X와 Y집단 각각에 대해 표준화된 열−비율의 차를 구한 다음, 그 차이를 곱하고 공간 근접성 행렬에서 주어지는 가중치를 또 곱하는 형태를 취하고 있다. 만일 공간 근접성 행렬이 행-표준화(row-standardization)되어 있다면, 이 수식의 의미는 보다 쉽게 이해될 수 있다. 즉, 한 공간단위에 대해, 이웃 공간단위와의 열-비율 차이의 교차-곱(cross-product)에 대한 가중 평균값을 산출하고, 모든 공간단위에 대해 그 가중 평균값을 합산하여 전체 통계량이 산출되는 것이다.

다음의 단계는 공간 연관성 측도와의 관련성을 고려하는 것이다. 앞에서 언급한 것처럼, 이변량 공간 연관성 측도는 많은 시사점을 주고 있는 데, 특히 이변량 Geary(bivariate Geary: BG)는 그 가능성이 높은 후 보이지만 아직까지 개발된 바 없다. 일변량 Geary의 수식을 변형하여 이변량 Geary를 구하면 다음의 수식으로 제시될 수 있다.

$$= \frac{n-1}{2\sum_i \sum_j v_{ij}} \cdot \frac{\sum_i \sum_j v_{ij}(x_i - x_j)(y_i - y_j)}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \qquad (8)$$

이변량 Geary의 통계학적인 속성에 대해서는 다른 논문에서 자세히 다루어져야 하지만, 두 가지 사항은 여기서 언급될 필요가 있다. 첫째, 기대값이 두 변수 간의 피어슨의 상관계수(Pearson's coefficient of correlation)로 주어진다. 이것은 공간적 분리가 두 집단의 열-비중 간의 상관관계에 의해 규정되어야 한다는 점을 감안할 때 매우 고무적인 사실이 아닐 수 없다. 둘째, 구조적으로 식(7)과 매우 유사한 형태를 취하고 있다. 이는 식(7)을 식(8)의 형태로 변환하기만 하면 공간 연관성 측도에서 이미 밝혀진 다양한 통계학적 결과를 이용할 수 있게 된다는 점을 의미한다. 식(7)을 식(8)의 구조에 맞추어 변형하여 최종적인 공간 분리성 측도(SSM)를 제시하면 다음과 같다.

$$SSM = \frac{n-1}{2\sum_i \sum_j v_{ij}} \cdot \frac{\sum_i \sum_j v_{ij}[(r_i^X - p_i) - (r_j^X - p_j)] \cdot [(r_i^Y - p_i) - (r_j^Y - p_j)]}{\sqrt{\sum_i (r_i^X - p_i)^2}\sqrt{\sum_i (r_i^Y - p_i)^2}} \qquad (9)$$

분자는 식(7)를 따라가지만 분모와 상수는 측도의 표준화를 위해 BG에서 사용된 것을 수용했다. 식(9)를 행렬의 형태로 표현하면 다음과 같다.

$$SSM = \frac{n-1}{n} \cdot \frac{(\mathbf{z}^X)^T(\mathbf{\Omega}-\mathbf{V}^S)\mathbf{z}^Y}{\mathbf{1}^T\mathbf{V1}} \tag{10}$$

이 때, **1**은 모든 요소가 1로 이루어진 $n\times1$벡터를 의미하며, $\mathbf{z}^X$와 $\mathbf{z}^Y$벡터의 각 요소는 다음과 같이 주어진다.

$$z_i^X = \frac{r_i^X - p_i}{\sqrt{\sum_i (r_i^X - p_i)^2/n}} \quad \text{그리고} \quad z_i^Y = \frac{r_i^Y - p_i}{\sqrt{\sum_i (r_i^Y - p_i)^2/n}} \tag{11}$$

또한 공간 근접성 행렬이 복잡한 형태로 주어져 있는데, 이러한 변환은 전체 수식을 이차형태(quadratic form)으로 전환하기 위해 필수적이며(Lee, 2004; 2007), 이후에서 논의될 유의성 검정법 개발에서도 요구되는 것이다. 좀 더 자세히 제시하면 다음과 같다.

$$\mathbf{\Omega}-\mathbf{V}^S = \begin{bmatrix} \sum_j v_{1j}^S - v_{11}^S & -v_{12}^S & \cdots & -v_{1n}^S \\ -v_{21}^S & \sum_j v_{2j}^S - v_{22}^S & \cdots & -v_{2n}^S \\ \vdots & \vdots & \ddots & \vdots \\ -v_{n1}^S & -v_{n2}^S & \cdots & \sum_j v_{nj}^S - v_{nn}^S \end{bmatrix} \tag{12}$$

$$\mathbf{V}^S = 0.5 \times [\mathbf{V} + \mathbf{V}^T]$$

3) 유의성 검정

유의성 검정은 '일반화된 랜덤화 검정법(generalized randomization test)'에 의해 이루어질 수 있다(Lee, 2004; 2007). 이 검정법은 공간 연관성 측도의 가설검정을 위해 제시된 것으로, 공간적 독립성이라는 귀무가설에 대한 가설검정을 행할 수 있게 해준다. 일반화된 랜덤화 검정법은 크게 '확장–만텔 검정(extended Mantel test)'과 '일반화된 벡터 랜덤화 검정(generalized vector randomization test)'으로 구성되고, 각 검정법은 측도의 종류와 전역적이냐 국지적이냐를 고려하면서 사용될 수 있다(Lee, 2004; 2007). 이 중 전역적 통계량의 유의성 검정에 사용될 확장-만텔 검정을 간단히 소개하고자 한다.

확장–만텔 검정은 Mantel(1967)의 시-공간 의존성(spatio-temporal dependence) 측도의 가설검정을 확장한 것으로, Heo and Gabriel(1998)의 연구에 바탕을 두고 있다. 이것은 만일 어떤 측도 혹은 통계량이 두 행렬의 케이스 별 곱의 합으로 표현될 수 있다면, 그 측도 혹은 통계량의 기대값과 분산은 일반화된 공식에 의해 구해질 수 있다는 것이다. 우선 측도는 다음의 형태로 정의될 수 있어야 한다.

$$\Gamma = \sum_i \sum_j p_{ij}\, q_{ij} = \sum_{i,j} (\mathbf{P} \circ \mathbf{Q}) = \mathrm{tr}(\mathbf{P}^T\mathbf{Q}) = \mathrm{tr}(\mathbf{P}\mathbf{Q}^T) \tag{13}$$

여기서 **P**는 표준화된 공간 근접성 행렬을 의미하고, **Q**는 특정한 형태의 수치적 유사성 행렬(numeric similarity matrix)을 의미한다. 즉, 전자는 공간단위 간의 위상 관계를 정의하고, 후자는 공간단위가 보유한 값들 간의 관계를 정의한다. 또 tr( )는 행렬의 대각선에 위치한 모든 값을 합산하는 행렬 오퍼레이션을 의미한다. 만일 측도가 위의 형태로 정의된다면, 이 측도의 기대값과 분산은 다음의 공식에 의해 주어진다.

$$\mathrm{E}(\Gamma) = \mathrm{E}(\Gamma^{\mathrm{off}}) + \mathrm{E}(\Gamma^{\mathrm{on}}) \tag{14}$$

$$\mathrm{Var}(\Gamma) = \mathrm{Var}(\Gamma^{\mathrm{off}}) + \mathrm{Var}(\Gamma^{\mathrm{on}}) + 2 \cdot \mathrm{Cov}(\Gamma^{\mathrm{off}}, \Gamma^{\mathrm{on}}) \tag{15}$$

식(14)와 식(15)를 보면, 기대값을 위해서는 두 개의 요소, 분산을 위해서는 세 개의 요소에 대한 정의가 필요하다는 것을 알 수 있는데, 각각의 요소를 정의하기 위해서는 또 다른 다수의 양(quantities)에 대한 정의가 필요하다(Lee, 2004 참조).

결국 중요한 것은 식(9)에서 제시된 공간 분리성 측도가 식(13)과 같은 형태로 변형될 수 있느냐 하는 점이다. 행렬의 형태로 표현된 식(10)을 이용하면, 공간 분리성 측도가 다음의 두 행렬의 케이스별 곱의 합으로 변형될 수 있음을 알 수 있다.

$$\mathbf{P} \equiv \frac{n-1}{n} \cdot \frac{(\mathbf{\Omega}-\mathbf{V}^S)}{\mathbf{1}^T\mathbf{V1}} \tag{16}$$

$$\mathbf{Q} \equiv \mathbf{z}_X \cdot (\mathbf{z}_Y)^T \tag{17}$$

여기서 $\mathbf{1}^T\mathbf{V1}$는 공간 근접성 행렬의 모든 요소를 합산한다는 것을 의미하며, $\mathbf{\Omega}-\mathbf{V}^S$는 식(12)에, $\mathbf{z}^X$와 $\mathbf{z}^Y$는 식(11)에 제시되어 있다. 이 두 행렬를 이용하면 공간 분리성 측도의 기대값과 분산을 구할 수 있는 수식을 얻을 수 있다.

분산을 구하는 과정은 매우 복잡하기 때문에 여기서는 기대값을 구하는 수식이 어떻게 도출되는지 만을 제시하고자 한다. 식(14)를 보다 자세히 표현하면 다음과 같다(Lee, 2004).

$$E(\Gamma)=E(\Gamma^{\text{off}})+E(\Gamma^{\text{on}})$$
$$=\frac{[(\mathbf{1}^T\mathbf{P}\mathbf{1})-\text{tr}(\mathbf{P})]\cdot(\mathbf{1}^T\mathbf{Q}\mathbf{1})-\text{tr}(\mathbf{Q})]}{n(n\text{-}1)}+\frac{\text{tr}(\mathbf{P})\cdot\text{tr}(\mathbf{Q})}{n} \quad (18)$$

식(16)과 (17)에서 규정된 두 행렬을 식(18)에 의거해 계산하면 결국 기대값은 다음과 같이 주어진다.

$$E(SSM)=\left[1-\frac{\text{tr}(\mathbf{V})}{\mathbf{1}^T\mathbf{V}\mathbf{1}}\right]\cdot r_{X,Y} \quad (19)$$

왜냐하면, $\mathbf{1}^T\mathbf{P}\mathbf{1}=0$, $\text{tr}(\mathbf{P})=\frac{n-1}{n}\cdot\left(1-\frac{\text{tr}(\mathbf{V})}{\mathbf{1}^T\mathbf{V}\mathbf{1}}\right)\cdot$ $\mathbf{1}^T\mathbf{Q}\mathbf{1}=0$, 그리고 $\text{tr}(\mathbf{Q})=n\cdot r_{X,Y}$로 단순화될 수 있기 때문이다. 이 때 $r_{X,Y}$는 두 변수 X와 Y간의 상관계수를 의미하고, X와 Y는 두 집단의 표준화된 열–비중을 의미한다.

$$X=\begin{bmatrix} r_1^X-p_1 \\ r_2^X-p_2 \\ \vdots \\ r_i^X-p_i \\ \vdots \\ r_n^X-p_n \end{bmatrix} \text{ 그리고 } Y=\begin{bmatrix} r_1^Y-p_1 \\ r_2^Y-p_2 \\ \vdots \\ r_i^Y-p_i \\ \vdots \\ r_n^Y-p_n \end{bmatrix} \quad (20)$$

그런데, 이변량 Geary의 경우 공간 근접성 행렬의 대각선에 0을 놓는 것이 일반적이므로 분자인 $\text{tr}(\mathbf{V})=0$이 되어 최종적인 기대값은 $r_{X,Y}$로 주어진다. 이러한 방식으로 기대값과 분산이 주어지면, z-값을 계산한 후, 표본분포(sampling distribution)가 정규분포를 따른다는 가정 하에 가설검정을 행하는 '정규 근사(normal approximation)'를 통해 유의성 검정이 이루어지게 된다.

4) 결과의 해석

공간 분리성 측도는 이변량 Geary의 수식에 기반을 둔 것이기 때문에, 우선 일변량 Geary를 해석하는 방식을 검토할 필요가 있다. 일변량 Geary는 0 이상의 값을 가지며, 기대값은 1이다. 이 때 측도값이 1 보다 크면, 즉 z-값이 0 보다 크면, 공간적 패턴이 음(-)의 공간적 의존성을 갖는다는 것을 의미하고, 반대로 측도값이 1 보다 작으면, 즉 z-값이 0 보다 작으면, 양(+)의 공간적 의존성을 갖는다는 것을 의미한다. 측도값이 0에 가까워지면 질수록 공간적 자기상관의 정도는 높아진다. 그러나 이변량 Geary는 이보다 훨씬 더 복잡한데, 이는 측도값과 기대값이 음수일수도 있고 양수일 수도 있기 때문이다. 따라서 공간 분리성 측도와 유의성 검정 결과를 해석하기 위해서는 다음의 몇 가지 개념적 명료화가 요구된다.

첫째, 우선적으로 공간 분리성 측도(혹은 기대값)가 양의 값을 갖는 다는 것은 두 인구 집단의 표준화된 열-비중 간에 양(+)적인 관련이 존재한다는 것을 의미하고, 음의 값을 갖는 다면 그 반대라는 것을 의미한다. 이 때 최소한 음적인 관계를 보일 때 두 인구 집단 간에 공간적 분리가 존재하는 것으로 판정하고자 한다. 공간 분리성 측도가 음의 값을 보인다면, 도시 전체적으로 두 집단이 공간적으로 분리되어 있다고 해석할 수 있고, 양의 값을 나타낸다면 두 집단의 거주공간이 주어진 공간단위 수준에서 중첩되어 나타나고 있다고 해석할 수 있다. 다시 말해, 측도가 음의 값을 갖는다는 것은 한 인구 집단이 지배적으로 거주하고 있는 공간단위의 수가 상대적으로 많다는 것을 의미한다. 상이지수는 모두 양의 값만을 가지기 때문에 분리/중첩의 기준점을 찾을 수 없다는 단점이 있었는데, 공간 분리성 측도는 그러한 점을 해소하고 있다. 이 때 주의할 것은 공간 분리성 측도의 부호는 단지 상이지수처럼 많은 공간단위에서 한 인구 집단이 지배적으로 거주하는 경향이 있는지 그렇지 않은지 만을 표시할 뿐 그러한 공간단위가 공간적으로 집중하고 있는지 그렇지 않은지, 즉 공간적 의존성의 정도에 대해서는 아무런 정보도 제공하고 있지 않다는 사실이다. 따라서 측도의 부호는 상이지수와 마찬가지로 비공간적인 의미의 거주지 분리/중첩의 상황만을 표시해 준다.

둘째, 공간 분리성 측도와 기대값과의 관계가 공간적 의존성의 정도를 나타낸다. 즉, 공간 분리성 측도가 기대값으로부터 0의 방향으로 향하면 양(+)적인 공간적 의존성이 나타나는 것이고, 반대로 기대값으로부터

그림 3. 공간 분리성 측도의 해석

0과 반대 방향으로 향하면 음(-)적인 공간적 의존성이 나타나는 것을 의미한다. 예를 들어 기대값, 즉 표준화된 열-비중간의 상관관계가 -0.5이고 측도값이 -0.3이라면 양적인 공간적 의존성이 나타난다는 것이고, 측도값이 -0.7이라면 음적인 공간적 의존성이 나타난다는 것을 의미한다. 거꾸로 기대값이 0.5이고 측도값이 0.3이라면 양적인 공간적 의존성이 나타난다는 것이고, 측도값이 0.7이라면 음적인 공간적 의존성이 나타난다는 것을 의미한다. 가설검정의 결과 제시되는 z-값은 이러한 측도값과 기대값의 관계를 표현해 준다. 즉, z-값이 양수라는 것은 측도값이 기대값 보다 크다는 것을 의미하고 음수라는 것은 기대값 보다 작다는 것을 의미한다. 더 나아가 z-값의 절대치는 공간적 의존성의 강도를 표시한다. 즉, z-값의 크기가 크다면(즉 유의확률이 작다면) 양(+)적이든 음(-)적이든 공간적 의존성이 크다는 것을 의미한다.

이 두 가지 기준을 결합하면 그림 3에서 보듯이 네 가지 경우가 나오는 데 각각을 해석하면 다음과 같다. 첫째, 공간 분리성 측도값이 양수이면서 z-값이 양수이면(중첩-혼재) 두 집단의 거주 공간이 중첩되는 경향이 있으면서 동시에 그 패턴의 공간적 의존성이 음(-)적이라는 것을 의미한다. 둘째, 측도값이 양수이면서 z-값이 음수이면(중첩-클러스터), 두 집단의 거주 공간이 중첩되는 경향이 있으면서 동시에 양(+)적인 공간적 의존성이 존재한다는 것을 의미하는데, 이것은 전체적으로 거주지 분리의 일반적 경향은 잘 나타나지 않지만 분리/중첩의 양상이 유사한 공간단위들이 공간적으로 집중되어 있다는 것을 의미한다. 셋째, 공간 분리성 측도값이 음수이면서 z-값이 양수인 경우는(분리-클러스터) 거주지 분리의 경향성이 존재하면서 유사한 국지적 경향성을 보이는 지역들이 공간적으로 집중되어 있다는 것을 의미한다. 따라서 공간적 분리 현상에 있어 가장 주목해야 하는 것이 바로 이 경우라 할 수 있다. 넷째, 측도값이 음수이면서 z-값이 음수인 경우는(분리-혼재) 공간적 분리의 경향성이 존재하지만 그 공간적 의존성의 방향은 음(-)적이라는 것을 의미한다. 단적으로 표현하면, 그림 1에서 왼쪽의 그림은 이 네 번째 범주에 포함되는 것이고 오른쪽 그림은 위의 세 번째 범주에 포함되는 것이다.

4. 한국 대도시에의 적용

위에서 개발된 공간 분리성 측도와 유의성 검정법을

우리나라 7대 대도시에 적용하여 사례연구를 수행하였다. 거주지 분화 현상을 살펴보기 위한 지표로 최종학력을 사용하였다. 데이터는 2000년 인구 센서스 원자료 CD에서 추출한 40~59세 인구의 최종학력 자료이다. 최종학력을 기준으로 고학력(4년제 대졸 이상), 중학력(중졸 이상에서 4년제 대졸 미만), 저학력(중졸 미만)의 세 인구 집단으로 구분하고 인구 집단간 거주지 분리의 정도를 측정하였다. 기본적인 공간단위로 행정동을 사용했는데, 현 행정구역에 포함되어 있는 읍과 면 지역은 조사대상에서 제외하였다. 이것은 오랜 시간에 걸쳐 도시 지역이었던 곳만을 대상으로 하는 것이 연구의 목적에 보다 부합한다고 판단하였기 때문이다. 따라서, 부산광역시의 기장군, 대구광역시의 달성군, 인천광역시의 강화군과 옹진군, 울산광역시의 울주군은 분석에서 제외되었다. 공간 분리성 측도에 대한 계산과 유의성 검정은 S-Plus에서 제공되는 S-언어를 통해 관련 스크립트를 저자가 작성하여 사용하였다.

도시별 일반 사항을 요약하면 표 1과 같다. 표에서 살펴보면, 고학력 집단의 비중이 가장 높은 곳은 서울(19.85%)이며, 가장 낮은 곳은 울산(8.88%)이고, 반대로 저학력 집단의 비중이 가장 높은 도시는 대구(21.38%)이며, 가장 낮은 도시는 서울(14.02%)임을 알 수 있다. 이것은 서울의 경우 고학력과 저학력 집단간 거주지 분리의 수준이 높을 것임을 시사하는 것이며, 반대로 부산, 대구, 인천, 울산의 경우는 상대적으로 그 양상이 뚜렷하지 않을 것임을 시사하고 있다.

각 도시 별 세 인구 집단 간 거주지 분리에 대해 공간 분리성 측도에 의한 측정과 유의성 검정을 시행한 결과가 표 2에 나타나 있다.

상이지수에 의한 거주지 분리를 살펴보면, 모든 도시에서 고학력-저학력 간 거주지 분리가 가장 현저한 것으로 나타났으며, 뒤를 이어 고학력-중학력 간, 중학력-저학력 간 순으로 나타났다. 특히 서울과 대전의 경우 완전한 균형을 이루기 위해 주민의 약 47%가 이동해야 하는 것으로 나타났다. 일반적으로 이 정도의 수치는 매우 높은 것으로 파악되기 때문에 상이지수에 의거할 때 우리나라 대도시에서는 학력 집단 간 거주지 분리, 특히 고학력 집단과 저학력 집단 간의 거주지 분리가 뚜렷이 나타나고 있는 것으로 평가할 수 있다. 그러나 앞에서 언급한 것처럼, 분리/중첩을 구분할 수 있는 기준이 상이지수에는 존재하지 않는다.

공간 분리성 측도에 의거하면, 3개 도시(서울, 부산, 대전)의 중학력−저학력 간을 제외한 모든 집단 간에서 거주지 분리가 나타나고 있다고 말 할 수 있다. 기대값을 기준으로 살펴보면, 모든 시가 고학력-저학력 간에서 -0.8이하의 값을 보이고 있고, 고학력-중학력 간에는 그 보다 낮은 수준의 음적인 상관관계를 보이고 있으며, 중학력-저학력 간에는 낮은 음적인 상관관계이거나 양적인 상관관계가 나타나고 있다. 특히 서울의 중학력-저학력간 기대값은 0.66으로 이 두 집단 간에는 높은 수준의 거주지 중첩이 존재하는 것으로 판단된다.

유의수준 0.05에서 볼 때, 고학력 집단과 저학력 집단 사이에서 공간적 거주지 분리가 발생한 곳은 서울, 대구, 대전, 광주이다. 특히 서울의 경우 매우 높은 z-

표 1. 도시 별 학력 집단 비중

도시	분석 행정동 개수	학력 집단 비중 (%)		
		고학력	중학력	저학력
서울특별시	522	19.85	66.13	14.02
부산광역시	216	11.03	69.80	19.18
대구광역시	129	12.84	65.77	21.38
인천광역시	114	10.11	72.14	17.74
광주광역시	83	16.56	64.90	18.54
대전광역시	76	17.42	62.64	19.69
울산광역시	46	8.88	73.30	17.82

표 2. 도시 별 학력 집단 간 거주지 분리에 대한 공간 분리성 측도와 유의성 검정 결과

도시	분류	고학력 – 저학력	고학력 – 중학력	중학력 – 저학력
서울특별시	상이지수	0.4665	0.3376	0.1607
	SSM	-0.4175	-0.5363	0.3080
	기대값	-0.8647	-0.9485	0.6609
	z-값	**16.2575**	**12.9071**	**-13.6157**
부산광역시	상이지수	0.4168	0.2838	0.1603
	SSM	-1.0046	-0.9470	0.5095
	기대값	-0.8759	-0.6938	0.2602
	z-값	-1.4699	**-2.6617**	**3.3506**
대구광역시	상이지수	0.4158	0.2659	0.1720
	SSM	-0.4331	-0.1305	-0.2531
	기대값	-0.8652	-0.2848	-0.2340
	z-값	**8.2409**	**3.7024**	-0.4614
인천광역시	상이지수	0.3805	0.2423	0.1539
	SSM	-0.5965	-0.3621	-0.0498
	기대값	-0.8575	-0.4258	-0.1003
	z-값	**4.5311**	1.1268	1.1120
광주광역시	상이지수	0.3711	0.2032	0.1912
	SSM	-0.7808	-0.4117	-0.0070
	기대값	-0.8749	-0.3995	-0.0943
	z-값	1.2904	-0.1716	1.5257
대전광역시	상이지수	0.4665	0.3188	0.1874
	SSM	-0.4590	-0.4392	0.1488
	기대값	-0.8406	-0.7874	0.3456
	z-값	**5.5301**	**4.6092**	**-3.3977**
울산광역시	상이지수	0.3630	0.2590	0.1401
	SSM	-0.7932	-0.4497	-0.0346
	기대값	-0.8469	-0.4444	-0.1000
	z-값	0.5417	-0.0632	0.8395

값(-16.26)을 보이고 있어, 고학력이 지배적인 동들이 공간적으로 집중되어 있을 뿐만 아니라 저학력이 지배적인 동들도 공간적으로 인접하는 경향이 강한 것으로 판단할 수 있다. 유의확률을 기준으로 할 때, 공간적 거주지 분리의 정도는 서울에 이어 대구, 대전, 인천의 순인 것으로 드러났다. 반면에, 부산, 광주, 울산은 통계적으로 유의한 만큼의 공간적 분리 현상을 보이고 있지 않은 것으로 분석되었다.

고학력 집단과 중학력 집단 간에 통계적으로 유의한 만큼의 공간적 분리를 나타낸 시는 서울, 대전, 대구, 부산이다. 특히 서울의 경우는 고학력–저학력 간에 버금가는 수준의 공간적 분리가 나타나고 있다. 가장 특징적인 것은 부산의 경우인데, 나머지 시들이 모두 세 번째 범주(분리–클러스터)에 포함되어 있는데 반해 부산은 네 번째 범주(분리–혼재)인 것으로 나타났다. 이것은 상당히 높은 수준의 거주지 분리에도 불구하고

고학력 집단이 지배적인 동과 중학력 집단이 지배적인 동들이 각자 클러스터를 형성하지 못하고 서로서로 혼재되어 분포하는 경향이 강하기 때문인 것으로 판단된다. 다시 말해서, 고학력 집단이 지배적인 동들 주변에 오히려 중학력 집단이 지배적인 동들이 위치하는 경우가 적지 않다는 것을 의미하는 것이다.

중학력 집단과 저학력 집단 간에 통계적으로 유의한 만큼의 공간적 분리를 나타낸 시는 서울, 대전, 부산이다. 서울과 대전의 경우는 두 번째 범주(중첩–클러스터)에 포함되는 데, 두 집단 간 거주지의 중첩도가 높으면서 그러한 특성을 보이는 동들이 공간적으로 집중되어 나타나고 있음을 알 수 있다. 특히 서울의 경우는 두 집단의 거주지가 중첩됨과 동시에 유사한 중첩 양상을 보이는 지역이 공간적으로 매우 집중되어 있는 것을 알 수 있다. 여기서도 부산은 매우 예외적인 사례로 나타나는 데 두 집단 간 거주지가 중첩되는 경향이 강하지만 음(-)적인 공간적 의존성이 나타나는 것이다. (중첩–혼재)거주지의 중첩도가 높은 동들 주변에 오히려 거주지 분리가 두드러진 동들이 위치하는 등의 상황이 시 전역에서 나타나고 있음을 함축하고 있다.

전체적으로 봤을 때, 서울과 대전은 거주지 분리의 매우 전형적인 패턴을 보여주고 있는 것으로 판단된다. 즉, 고학력-저학력, 고학력-중학력 간에서는 강한 거주지 분리 경향과 함께 정(+)적인 공간적 의존성을 보여주고 있고, 중학력-저학력 간에는 거주지 중첩의 경향과 함께 역시 정(+)적인 공간적 의존성을 보여주고 있다. 부산의 경우는 거주지 분리/중첩의 경향성은 서울, 대전과 동일하지만 정(+)적인 공간적 의존성의 정도는 현저히 떨어지는 것으로 드러났다. 인천은 고학력-저학력 집단 사이에서만 통계적으로 유의한 공간적 분리가 관찰되었고, 광주와 울산에서는 통계적으로 유의한 공간적 분리의 경향이 발견되지 않았다.

5. 요약 및 결론

이 연구는 거주지 분리의 공간적 차원을 충실히 반영할 수 있는 새로운 공간 분리성 측도와 그것에 대한 통계적 검정법을 개발함으로써 집단 간 거주지 분리를 측정하는 새로운 도구를 제시하고자 하였으며, 그것의 유용성을 우리나라 대도시에 대한 사례연구를 통해 예증하려 하였다. 중요한 결과는 다음과 같다. 첫째, 공간 연관성 측도와 공간적 카이–스퀘어 통계량을 통합함으로써 새로운 공간 분리성 측도를 개발하였다. 이 측도는 개별 공간 단위 내에서의 거주지 분리 양상뿐만 아니라 그러한 양상의 공간적 의존성도 함께 고려하고 있다. 둘째, 일반화된 랜덤화 검정법을 이용해 새로이 개발된 공간 분리성 측도에 대한 유의성 검정법을 제시하였다. 셋째, 공간 분리성 측도와 유의성 검정법을 우리나라 7대 대도시의 학력 집단 간 거주지 분화에 적용한 결과 높은 유용성이 있는 것으로 드러났다.

이 연구는 공간통계학의 최근 흐름을 반영하고 있는 것이다. 단순한 지표 혹은 통계량에서부터 복잡한 다변량 분석에 이르기까지 모든 종류의 통계학적 절차들은 데이터의 '공간성'을 고려함으로써 변형될 필요가 있다. 변동계수를 보완하기 위해 일변량 공간 연관성 측도가 필요하고, 상관계수를 보완하기 위해 이변량 공간 연관성 측도가 필요하고, 회귀분석을 보완하기 위해 공간적 회귀분석 기법이 필요한 것처럼, 상이지수를 보완하기 위해 공간 분리성 측도가 필요한 것이다. 본 연구는 주된 의의는 바로 여기에 있다.

이 연구의 가장 중요한 향후 과제는 제시된 전역적 공간 분리성 측도로부터 국지적 측도를 이끌어 내고, 그것을 바탕으로 한 다양한 탐색적 공간데이터분석(ESDA: exploratory spatial data analysis) 기법을 제시하는 것이다. 이 논문에서 제시된 측도는 본질적으로 전역적인 것이기 때문에 거주지 분화 패턴 전체에서 드러나는 공간적 의존성의 평균적인 경향성만을 고려하고 있다. 이에 반해 국지적 측도는 전역적 측도에 대한 각 공간단위의 상대적 중요성을 측정하는 것으로, 평균적인 경향성에 대한 국지적 공간단위의 조응 혹은 편기의 정도를 보여줌으로써, 공간적 이질성(spatial heterogeneity) 연구의 중요한 수단이 된다(Anselin, 1995; Getis and Ord, 1996). 전역적 공간 분리성 측도를 통해 드러난 거주지 분리의 일반적 양상이 도시 공간의 모든 부분에서 동일하게 구현될 것을 가정할 수 없기 때문에, 국지적 공간 분리성 측도를 이용하여 도시 공간의 어느 부분이 전역적 경향성에

조응하고 있으며, 어느 부분이 전역적 경향성에 반하고 있는지를 탐색할 필요성이 있는 것이다. 국지적 공간 분리성 측도를 이용한 탐색적 공간데이터분석 기법은 공간적 클러스터(spatial clusters), 공간적 특이점(spatial outliers), 공간적 체제(spatial regimes)를 확인할 수 있게 해주어야 하며(Anselin, 1995), 이를 위해서는 국지적 측도의 추출, 유의성 검정법의 개발, 그리고 관련 시각화 방법의 개발을 포함하는 전면적인 후속 연구가 요구되는 것이다.

文獻

최은영, 2004, 서울의 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.

Brown, L. A. and Chung, S.-Y., 2006, Spatial segregation, segregation indices and the geographical perspective, *Population, Space and Place*, 12(2), 125-143.

Chakravorty, S., 1996, A measurement of spatial disparity: the case of income inequality, *Urban Studies*, 33(9), 1671-1686.

Dawkins, C. J., 2004, Measuring the spatial pattern of residential segregation, *Urban Studies*, 41(4), 833-851.

Duncan, O. D. and Duncan, B., 1955, A methodological analysis of segregation indexes, *American Sociological Review*, 20(2), 210-217.

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.

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.

Heo, M. and Gabriel, K. R., 1998, A permutation test of association between configurations by mean of the RV coefficient, *Communications in Statistics: Simulation and Computation*, 27, 843-856.

Jakubs, J. F., 1981, A distance-based segregation index, *Socio-Economic Planning Sciences*, 15(3), 129-131.

Lee, C.-M. and Culhane, D. P., 1998, A perimeter-based clustering index for measuring spatial segregation: a cognitive GIS approach, *Environment and Planning B: Planning and Design*, 25(3), 327-343.

Lee, S.-I., 2001, 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., 2004, 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., 2007, A generalized randomization approach to local measures of spatial association, *Geographical Analysis*, under revision.

Maly, M. T., 2000, The neighborhood diversity index: a complementary measure of racial residential settlement, *Journal of Urban Affairs*, 22(1), 37-47.

Mantel, N., 1967, The detection of disease clustering and a generalized regression approach, *Cancer Research*, 27(2), 209-220.

Massey, D. S. and Denton, N. A., 1988, The dimensions of residential segregation, *Social Forces*, 67(2), 281-315.

Morgan, B. S., 1983, An alternative approach to the development of a distance-based measure of racial segregation, *American Journal of Sociology*, 88(6), 1237-1249.

Morrill, R. L., 1991, On the measure of geographic segregation, *Geography Research Forum*, 11, 25-36.

Oden, 1995, Adjusting Moran's *I* for population density, *Statistics in Medicine*, 14(1), 17-26.

Ransom, M. R., 2000, Sampling distributions of segregation indexes, *Sociological Methods & Research*, 28(4), 454-475.

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.

Tango, T., 1995, A class of tests for detecting 'general' and 'focused' clustering of rare diseases, *Statistics in Medicine*, 14(21-22), 2323-2334.

Wartenberg, D., 1985, Multivariate spatial correlation: a model for exploratory geographical analysis, *Geographical Analysis*, 17(4), 263-283.

Waldorf, B. S., 1993, Segregation in urban space: a new measurement approach, *Urban Studies*, 30(7), 1151-1164.

Waller, L. A. and Gotway, C. A., 2004, *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, Hoboken.

White, M. J., 1983, The measurement of spatial segregation, *American Journal of Sociology*, 88(5), 1008-1018.

White, M. J., 1986, Segregation and diversity: measures in population distribution, *Population Index*, 52(2), 198-221.

Wong, D. W. S., 1993, Spatial indexes of segregation, *Urban Studies*, 30(3), 559-572.

Wong, D. W. S., 1997, Spatial dependency of segregation indices, *The Canadian Geographer*, 41(2), 128-136.

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.

Wu, X. B. and Sui, D. Z., 2001, An initial exploration of a lacunarity-based segregation measure, *Environment and Planning B: Planning and Design*, 28(3), 433-446.

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 07. 09. 06.
최종접수일 07. 09. 12.