

격자자료분석을 위한 이웃정보시스템의 비교*

이강석¹⁾ 신기일²⁾

요약

최근 공간통계를 이용한 많은 연구가 진행되고 있고 공간통계학을 접목한 소지역 추정(small area estimation) 방법이 좋은 결과를 주고 있는 것으로 알려져 있다. 소지역 추정에 사용되는 격자자료(lattice data) 분석에서 이웃정보를 정의하는 것은 자료 분석의 성패를 결정짓는 매우 중요한 부분이다. 그러나 기존에 사용된 대부분의 이웃정보시스템은 경계선을 공유할 때 이웃으로 정하는 방법을 사용하고 있다. 이에 본 논문에서는 경계선 공유를 이용한 이웃정보시스템 뿐 아니라 다른 여러 이웃정보시스템을 구하는 방법을 설명하고 2001년 경제활동자료를 이용하여 이 시스템들을 비교하였다.

주요용어: 가중치, 이웃 정보, 공간상관관계, 조건부자기회귀공간모형.

1. 서론

공간통계자료는 크게 지리통계자료(geostatistical data)와 격자자료(lattice data)로 나누어진다. 격자자료 분석에서 가장 중요한 분석 단계는 각 자료의 공간상관관계를 규명하기 위한 이웃정보를 정의하고 이를 이용하여 이웃정보시스템을 구축하는 것이다. 이러한 이웃정보시스템은 분석할 자료의 특성에 따라 다르게 정의된다. 이러한 이유로 Cressie (1993)는 이웃정보를 정의하는 여러 가지 방법을 소개하였다. 어떤 이웃정보시스템을 사용하였느냐에 따라 그 분석결과의 우수성이 달라진다. 따라서 최적의 이웃정보를 구하는 것은 매우 중요하다. 여러 분야에서 공간통계를 이용한 연구가 활발히 진행되고 있다. 그 중에서 소지역추정에 공간통계학이 이용되고 있으며 Rao (2003)는 격자자료 분석에서 일반적으로 사용되는 조건부자기회귀공간모형(CAR: Conditional autoregressive spatial model) 모형을 접목한 소지역 추정모형을 설명하였다. 국내에서도 신기일과 이상은 (2003), 황희진과 신기일 (2008) 등과 같이 공간통계학을 이용한 소지역 추정에 관한 연구가 진행되고 있다. 그러나 이들이 사용한 이웃정보시스템은 매우 간단하다. 즉 경계선을 공유하고 있는 두 지역을 이웃으로 정한 이웃정보시스템을 사용하였다. 이 방법은 매우 간단하면서도 쉬운 방법으로 소지역이 정해지면 육안으로도 식별이 가능하다. 그러나 이 방법 이외의 이웃정보시스템 구축은 다른 기법을 이용하여야 가능하다. 이웃정보시스템을 구축하기 위한 가장 보편적인 방법은 GIS(지리정보시스템)을 이용하는 것이다. 최근 지리정보시스템을 이용한 많은 일들이 성공리에 이루어지고 있다. 공

* 이 연구는 2008년도 한국외국어대학교 교내연구비에 의해 수행되었음.

1) (449-791) 경기도 용인시 모현면 왕산리 산 89, 한국외국어대학교 자연과학대학 통계학과, 석사과정.

2) (449-791) 교신저자. 경기도 용인시 모현면 왕산리 산 89, 한국외국어대학교 자연과학대학 정보통계학과, 교수.

E-mail: keyshin@hufs.ac.kr

간통계에서도 지리정보시스템은 유용하게 사용되고 있고 지리정보시스템을 이용할 경우 매우 다양한 이웃정보시스템을 만들 수 있다. 본 논문에서는 지리정보시스템을 이용하여 이웃정보시스템을 결정하는 방법을 설명하였다. 또한 지리정보시스템에서 얻어진 기초 자료를 기반으로 새로운 이웃정보시스템을 구축하였다. 즉 지리정보시스템에서 얻어진 이웃정보를 다시 가공하여 분석할 자료에 맞는 이웃정보시스템을 구축하였다. 본 논문에서는 자료에 맞는 최적의 이웃을 정하는 방법을 설명하였으며 실제 한국에서 사용 가능한 이웃정보시스템 구축 방법을 제공하였다. 즉 이렇게 만들어진 다양한 이웃정보시스템을 공간통계학에서 가장 기본적으로 사용되는 Moran's I 통계량을 사용하여 비교하였다. Moran's I 통계량은 공간상관관계를 나타내는 척도로 그 값이 양이면 양의 공간상관관계를 음이면 음의 공간상관관계를 나타낸다. 일반적으로 높은 공간상관관계를 갖고 있는 이웃정보시스템은 좋은 분석결과를 줄 가능성이 매우 높다. 따라서 Moran's I 값을 기준으로 이웃정보시스템을 결정하는 것은 타당하다. 본 논문에서 사용한 자료는 2001년 4월 경제활동인구조사에서 전국(제주도 제외) 29,899 가구에서 얻어진 각 가구의 실업자 수이다. 이 자료는 신기일과 이상은 (2003)의 논문 및 김정오와 신기일 (2006)의 논문에서 사용된 같은 자료이다. 그리고 지리정보시스템 자료는 2001년도 행정구역기준의 각 시군구청간의 거리와 경계면 길이 그리고 시군구의 위치를 이용하였다. 본 논문의 구성은 다음과 같다. 먼저 2장에서 이웃정보를 구하는 여러 가지 방법을 설명하였다. 3장에서는 지리정보시스템을 이용하여 얻어진 이웃정보를 설명하였다. 4장에서는 경제활동인구 자료를 이웃정보시스템에 적용하여 얻은 Moran's I 값을 사용하여 그 결과를 비교하였다. 그리고 5장에 결론이 있다.

2. 이웃정보 구하는 방법

이웃 지역을 결정하는 방법은 각 조사 지역간의 공간적 위치와 거리를 기준으로 결정하는 방법과 각 지역간의 상호 교류 내역을 기준으로 정하는 방법이 있으나 본 논문에서는 거리를 기준으로 이웃정보시스템을 정하는 방법을 살펴보았다. 위치와 거리 기준은 크게 시군구청간의 거리 기준과 지역간 경계 공유 기준으로 나누어 진다. 여기에 가중치 개념이 추가로 도입 되기 때문에 같은 형태의 이웃으로 정해진 경우라도 가중치가 다르면 다른 이웃정보시스템을 갖게 된다.

2.1. 거리 기준

정해진 거리 안에 조사지역이 있는 경우 이웃(neighbor)으로 정한다. 이렇게 되면 어떤 지역에 해당되는 이웃의 수는 정해진 거리에 따라 달라지게 된다. 예를 들어 경기도 수원시의 경우 경계를 같이 하는 시군구는 안산, 의왕, 화성, 용인 등 4개 시이다. 각 시군구청간의 거리를 조사해 보면 안산시 18,650m, 의왕시 10,492m, 용인시 13,512m 그리고 화성시 18,936m이다. 만약 이웃이 되는 거리의 제한을 20km라 하면 모든 위의 4개 시 모두가 이웃이 된다. 그러나 15km로 제한하면 의왕시와 용인시만이 이웃이 된다. 다음으로 많이 사용하는 방법이 최근점 이웃(nearest neighbor)이다. 먼저 각 조사 시군구에서 가장 가까운 k 개(미리 정함)의 이웃 지역을 이웃으로 정하면 된다. 예를 들어 $k = 3$ 이 결정되면 수원시의 이웃은 안

산시, 의왕시 그리고 용인시가 된다.

2.2. 경계 기준

현재 가장 간단히 이웃을 정할 수 있는 방법은 지도상에서 같은 경계를 갖는 지역을 이웃으로 정하는 것이다. 그리고 경계를 공유하면 단순히 이웃으로 정하고 가중치는 모두 “1”로 가정한다. 가중치를 구하는 다른 방법은 경계의 길이, 각 시군구청간의 거리 등을 이용하여 가중치를 구하는 것이다. 이러한 가중치는 경계는 공유하지만 그 가중치가 다르기 때문에 각각 다른 이웃정보시스템을 갖는 것으로 생각해야 한다. 물론 가중치를 사용하였을 때 모든 자료에서 항상 좋은 결과를 주는것은 아니다. 다음 절에서 경계를 공유할 때 가중치를 구하는 방법을 설명하였다.

2.3. 가중값을 사용한 경우

2.1절과 2.2절은 거리에 상관없이 또는 경계선의 길이나 시군구청간의 거리에 상관없이 이웃 지역이 같은 영향을 준다는 가정 하에서 사용하는 이웃정보이다. 이는 거리가 멀면 영향력이 약화되는 일반적인 통념을 모형에 넣고 있지 않았다. 따라서 거리가 멀어지면 상관관계가 약화되는 모형이 필요하게 된다. 다음은 Cressie (1993)에 나와있는 가중치를 구하는 방법이다.

$$w_{ij} = \begin{cases} w_{ij}, & \text{만약 } j \text{와 } i \text{가 이웃이면,} \\ 0, & \text{그렇지 않으면,} \\ 0, & \text{만약 } i = j, \end{cases}$$

여기서 w_{ij} 는 다음의 여러 가지 방법으로 정의 될 수 있으며 $\sum_j w_{ij} = 1$ 을 만족하도록 하였다.

- 1) $w_{ij} = d_{ij}^{-\gamma}$, $\gamma > 0$, d_{ij} 는 조사 지역 i 와 j 의 거리로 본 논문에서는 각 시군청간의 거리를 나타낸다. 거리가 멀어지면 멀어질수록 그 가중값은 줄어든다. 이는 거리가 멀어지면 일반적으로 상관관계가 줄어드는 것을 모형화한 방법이다.
- 2) $w_{ij} = (l_{ij}/L_i)^\tau$, $\tau > 0$, 여기서 $L_i = \sum_j l_{ij}$ 이고 l_{ij} 는 i 번째 지역과 j 번째 지역의 공통 경계선의 길이이다. 따라서 공통 경계선이 길어지면 영향을 많이 주고 역으로 경계선이 짧아지면 그 영향력이 작아지는 것을 모형화한 것이다.
- 3) $w_{ij} = (l_{ij}/L_i)^\tau / d_{ij}^\gamma$. 두 조사 지역간의 거리와 공통 경계선의 길이를 함께 사용하여 모형화한 것이다.

3. GIS를 이용한 이웃 정보

사용된 자료는 2001년도 4월 자료이며 2001년도 행정구역을 기준으로 각 시군구의 위치와 경계면 길이, 그리고 시군구청간 거리 등의 자료를 이용하여 이웃 정보를 정하였다. 본 논문에서 사용할 이웃정보는 다음과 같다.

3.1. 경계 공유

경계를 공유할 때 이웃이라 정한다. 사용한 가중치는 다음과 같다. 여기서 w_{ij} 는 i 와 j 이웃지역의 가중치이며, k 는 이웃의 수, d_{ij} 는 이웃의 거리이다. 그리고 $d_{ij}^{-\gamma}$ 에서 $\gamma = 0, 1, 2, 5, 10, 15, 20, 30$ 을 사용하였다. 여기서 이웃의 수 k 는 지역에 따라 다르다.

| 방법 | 1 | 2 | 3 | 4 | 5 |
|----------|---|-------|-----------------------|-------------------------|--------------------|
| w_{ij} | 1 | $1/k$ | $(l_{ik}/L_i)^\gamma$ | $(l_{ik}/L_i)/d_{ij}^2$ | $d_{ij}^{-\gamma}$ |

3.2. 가장 가까운 시군구 이웃

가장 가까운 시군구 k 개를 이웃으로 정한다. 사용한 가중치는 다음과 같다.

| 방법 | 1 | 2 | 3 |
|----------|---|-------|--------------------|
| w_{ij} | 1 | $1/k$ | $d_{ij}^{-\gamma}$ |

본 논문에서는 이웃의 수 $k = 2, 3, 4$ 를 사용하였다. 여기서 w_{ij} , d_{ij} 그리고 γ 는 3.1절에서 정의한 것과 같다.

3.3. 거리 기준

거리는 각 시군구 청사간의 거리를 기준으로 정하였다. 각 시도마다 기준 거리를 다르게 적용하였다. 그렇지 않을 경우 어떤 시도는 그 시도내의 모든 시군구를 이웃으로 삼게 된다. 따라서 서울을 비롯한 다른 광역시들은 10km 이내, 경기도를 제외한 다른 도는 광역시들에 비해 시군들이 크고 시군간의 거리가 멀기 때문에 30km 이내의 지역을 이웃으로 정하였다. 그리고 경기도는 다른 도와 달리 광역시들 처럼 작은 시군들도 많이 존재하여 기준거리를 20km 이내로 정한후 이웃으로 잡았다. 이때 사용한 가중치는 다음과 같다.

| 방법 | 1 | 2 | 3 |
|----------|---|-------|--------------------|
| w_{ij} | 1 | $1/k$ | $d_{ij}^{-\gamma}$ |

여기서 이웃의 수 k 는 지역에 따라 다르며 w_{ij} , d_{ij} 그리고 γ 는 3.1절에서 정의한 것과 같다.

4. 응용 사례

본 논문에서 사용한 자료는 2001년 4월 경제활동인구조사 자료의 실업자수이며 이는 신기일과 이상은 (2003)에서 사용한 자료이기도 하다. 모형 설정에 앞서 먼저 자료를 조사가구 수 n_i 와 실업자 수 X_i 를 다음의 Freeman-Tukey (1950) 변환을 이용하여 변환하였다. 이는 신기일과 이상은 (2003)에서 사용한 변환방법이다.

$$Y_i = \left(1000 \times \frac{X_i}{n_i}\right)^{\frac{1}{2}} + \left(1000 \times \frac{X_i + 1}{n_i}\right)^{\frac{1}{2}}.$$

변환된 시군구별 자료가 공간상관관계를 검정하는데 사용되었다.

표 4.1: 경계 공유 이웃의 가중치별 Moran's I

| 통계량 | 가중치 | 모수 γ | 추정량 | 표준오차 | |
|-----------|-------------------------|-------------|----------|----------|----------|
| Moran's I | $w_{ij}=1$ | | 0.372 | 0.051 | |
| | $w_{ij}=\frac{1}{k}$ | | 0.302 | 0.054 | |
| | $(l_{ik}/L_i)^\tau$ | | 0 | 0.302 | 0.054 |
| | | | 1 | 0.321 | 0.060 |
| | | | 2 | 0.337 | 0.066 |
| | | | \vdots | \vdots | \vdots |
| | | | 30 | 0.370 | 0.085 |
| | $(l_{ik}/L_i)/d_{ij}^2$ | | 0.316 | 0.068 | |
| | $d_{ij}^{-\gamma}$ | | 0 | 0.302 | 0.054 |
| | | | 1 | 0.300 | 0.058 |
| | | | 2 | 0.296 | 0.063 |
| | | | \vdots | \vdots | \vdots |
| | | | 30 | 0.264 | 0.086 |

Moran's I 는 공간 통계학에서 공간상관관계를 검정하기 위한 검정 통계량으로 가장 많이 사용되고 있으며 정의는 다음과 같다.

$$\text{Moran's I} = \frac{n \sum_{i=1}^n \sum_{j=1}^n \delta_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\left(\sum_{i \neq j} \right) \delta_{ij} \sum_{i=1}^n (Z_i - \bar{Z})^2}$$

여기서 δ_{ij} 는 이웃이면 1이고 이웃이 아니면 0인 지시변수이다. 울릉군과 같은 도서 군과 자료가 없는 시군구는 분석에서 제외되어 사용된 자료 수 $n = 210$ 이다. 가중치를 사용한 일반화된 Moran's I 통계량 공식은 Cressie (1993) 또는 Cliff와 Ord (1981)을 참조하고 계산 방법은 Kaluzny 등 (1998)을 참조하기 바란다.

본 논문에서는 Moran's I를 비교통계량으로하여 3절에서 제안된 여러 방법을 비교하였다.

4.1. 경계 공유

3.1절에서 사용한 “경계를 공유하면 이웃”으로 정한 이웃 정보와 Freeman-Tukey (1950) 변환된 실업자수를 이용하여 Moran's I를 구하였다.

표 4.1 을 살펴보면 경계를 공유하면 이웃이라 정할때 $w_{ij}=1$ 인 경우의 Moran's I가 가장 큰 것을 알 수 있다. 이 경우, 신기일과 이상은 (2003)의 논문에서 구한 $w_{ij}=1$ 일 때의 Moran's I 값과 거의 일치한다. 그림 4.1과 4.2는 표 4.1 의 $(l_{ik}/L_i)^\tau$ 과 $d_{ij}^{-\gamma}$ 의 결과를 그림으로 나타낸 것이다.

$(l_{ik}/L_i)^\tau$ 의 Moran's I 값은 표 4.1에서 모수 τ 의 값이 커질수록 Moran's I 값이 증가하는 것을 알 수 있다. 그러나 모수 τ 가 5를 지나면서 더 이상 증가하지 않고 있다.

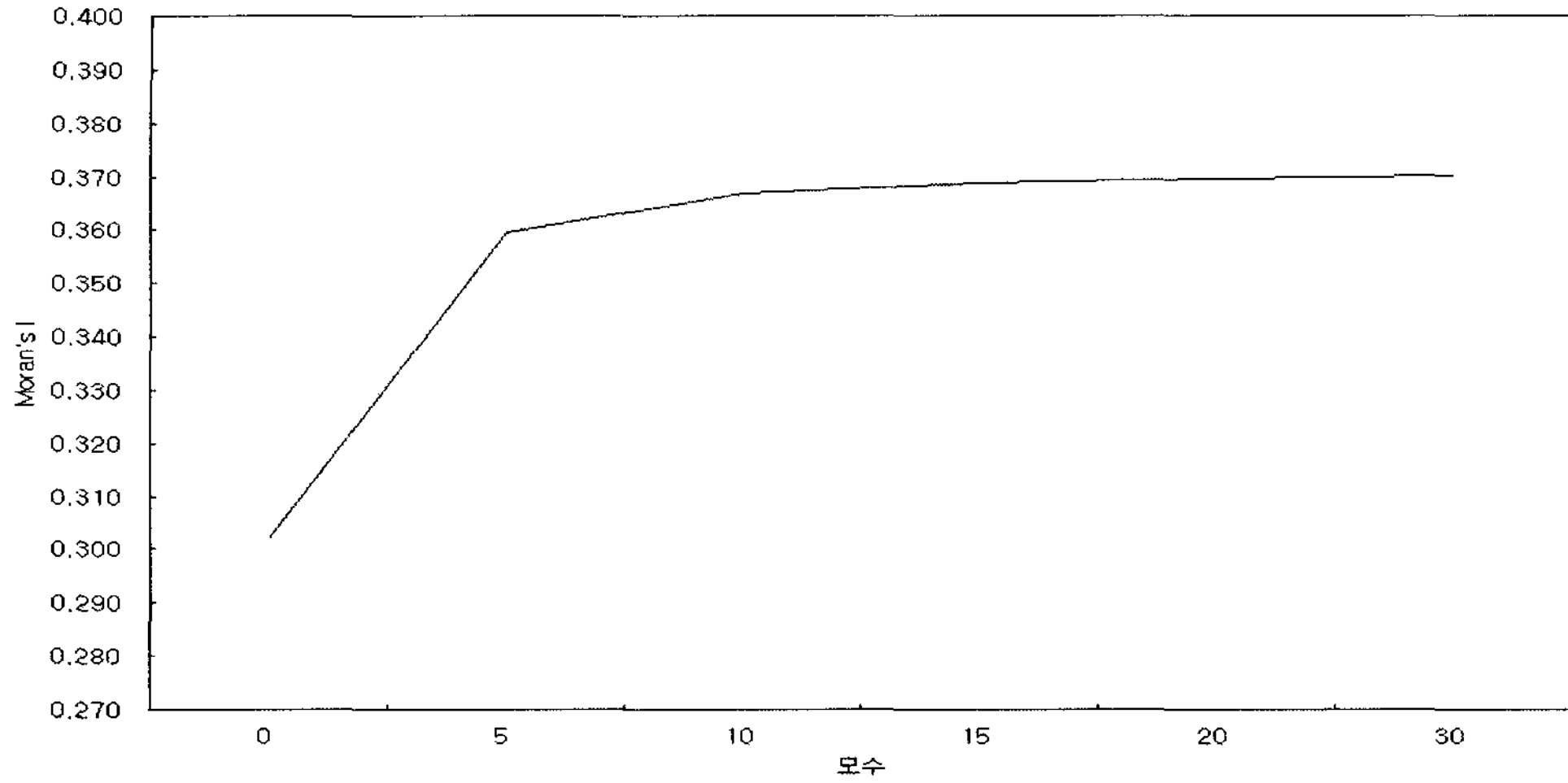


그림 4.1: $(l_{ik}/L_i)^\tau$ 의 모수 τ 에 따른 Moran's I 값 비교

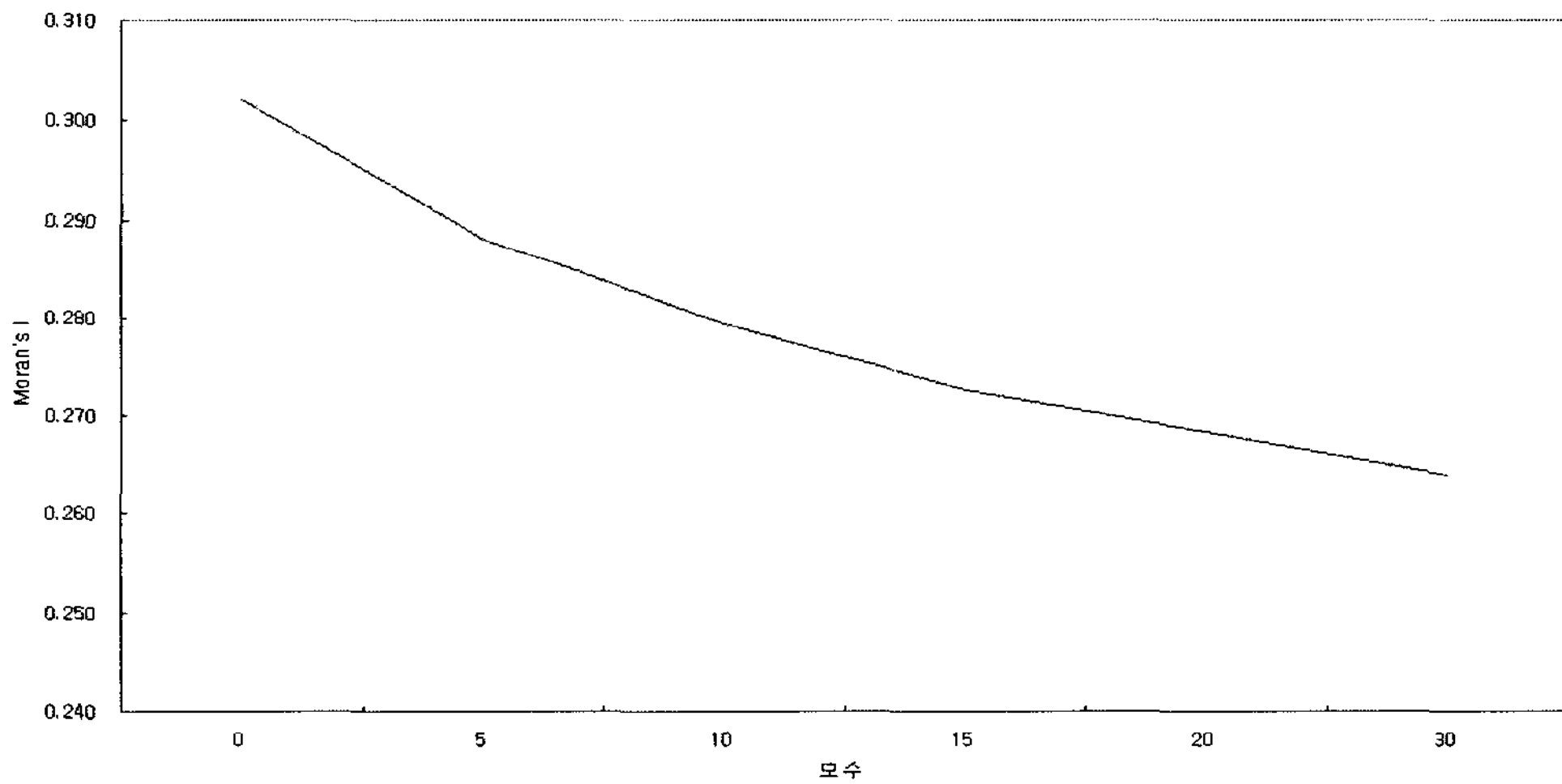


그림 4.2: $d_{ij}^{-\gamma}$ 의 모수 γ 에 따른 Moran's I 값 비교

그림 4.2를 살펴보면 $d_{ij}^{-\gamma}$ 의 Moran's I 값은 모수 γ 의 값이 커질수록 작아지는 것을 확인할 수 있다. 다음으로 두 그림을 비교해 보면 그림 4.1는 모수 τ 가 충분히 커지면 가장 긴 경계선을 가진 이웃의 가중이 “1”이 되고 다른 이웃의 가중은 “0”이 된다. 반면 그림 4.2에서 모수 γ 가 충분히 커지게 되면 가장 가까운 이웃의 가중은 “1”이 되고 다른 이웃의 가중은 “0”이 된다. 따라서 하나의 이웃을 사용한다면 경계선의 길이가 가장 긴 것을 사용하는 것이 가장 거리가 짧은 것을 사용하는 것보다 더 좋은 결과를 주고 있음을 확인할 수 있다.

표 4.2: 시군구 k 개 이웃의 가중치별 Moran's I

| 시군구수(k) | 가중치 | 모수 γ | 추정량 | 표준오차 |
|-------------|--------------------|-------------|----------|----------|
| $k=2$ | $w_{ij}=1$ | | 0.310 | 0.061 |
| | $w_{ij}=1/k$ | | 0.310 | 0.061 |
| | $d_{ij}^{-\gamma}$ | 0 | 0.310 | 0.061 |
| | | 1 | 0.302 | 0.065 |
| | | 2 | 0.297 | 0.068 |
| | | \vdots | \vdots | \vdots |
| 30 | | 0.265 | 0.084 | |
| $k=3$ | $w_{ij}=1$ | | 0.324 | 0.051 |
| | $w_{ij}=1/k$ | | 0.324 | 0.051 |
| | $d_{ij}^{-\gamma}$ | 0 | 0.324 | 0.051 |
| | | 1 | 0.314 | 0.056 |
| | | 2 | 0.306 | 0.062 |
| | | \vdots | \vdots | \vdots |
| 30 | | 0.264 | 0.085 | |
| $k=4$ | $w_{ij}=1$ | | 0.282 | 0.045 |
| | $w_k=1/k$ | | 0.282 | 0.045 |
| | $d_{ij}^{-\gamma}$ | 0 | 0.282 | 0.045 |
| | | 1 | 0.284 | 0.050 |
| | | 2 | 0.284 | 0.058 |
| | | \vdots | \vdots | \vdots |
| 30 | | 0.265 | 0.085 | |

4.2. 가장 가까운 시군구 이웃

전술하였듯이 본 논문에서는 가장 가까운 이웃 시군구 수를 $k = 2, 3, 4$ 로 정하여 모의실험을 시행하였다. 이에 관한 결과가 표 4.2에 나와 있다. 이 절에서도 d_{ij} 는 두 이웃간의 거리를 나타내며 거리의 변화에 따른 Moran's I의 변화를 보기 위해 모수 γ 를 도입하였다. 또한 γ 에 따른 Moran's I의 변화를 보기위해 이를 그림 4.3에 나타내었다.

그림 4.3을 살펴보면 이웃의 수 k 와 상관없이 모든 k 에서 γ 가 증가함에 따라 Moran's I 값은 감소하는 것을 알 수 있다. 또한 γ 가 10 이상이 되면 Moran's I 값이 모두 약 0.265에 접근하는 것을 확인 할 수 있다. 결론적으로 주어진 자료에서는 $k = 3$ 이고 $w_k = 1$ 또는 $1/k$ 일 때 가장 큰 Moran's I를 주는 것으로 나타났다.

4.3. 거리 기준

3.3절에서 정한 거리 기준에 의해 모의실험을 실시하였다. 즉 특별시와 광역시는 10km 이내, 경기도 20km이내, 나머지 다른 도는 30km 이내에 시군구청이 있으면 이를 이웃으로 정하였다. 이렇게 정해진 이웃정보시스템에 의해 구해진 Moran's I 값을 표 4.3과 그림 4.4에 정리하였다.

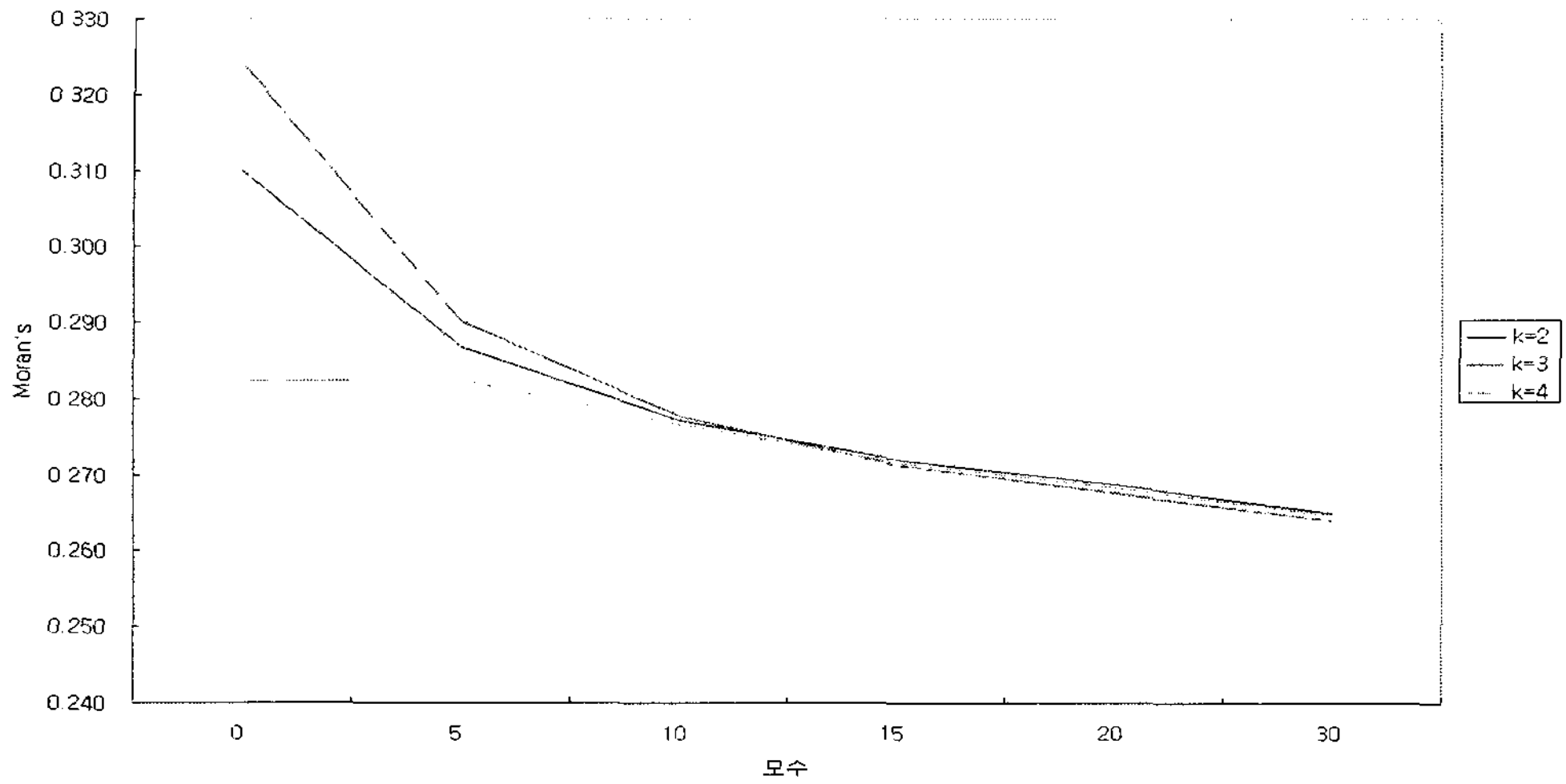


그림 4.3: $d_{ij}^{-\gamma}$ 의 모수 γ 에 따른 Moran's I 값 비교

표 4.3: 거리 기준 이웃의 가중치별 Moran's I

| 통계량 | 가중치 | 모수 γ | 추정량 | 표준오차 | |
|-----------|--------------------|-------------|-------|-------|-------|
| Moran's I | $w_{ij}=1$ | | 0.336 | 0.053 | |
| | $w_{ij}=1/k$ | | 0.320 | 0.054 | |
| | $d_{ij}^{-\gamma}$ | | 0 | 0.320 | 0.054 |
| | | | 1 | 0.309 | 0.058 |
| | | | 2 | 0.300 | 0.063 |
| | | | ⋮ | ⋮ | ⋮ |
| | | 30 | 0.265 | 0.085 | |

그림 4.4를 살펴보면 γ 가 증가함에 따라서 $d_{ij}^{-\gamma}$ 의 Moran's I 값은 감소하는 것을 확인 할 수 있다. 이러한 현상은 $d_{ij}^{-\gamma}$ 에서 γ 가 커지면 가장 가까운 이웃에 가중치 “1”을 주기 문에 나타나는 공통된 현상이다.

4.4. 응용사례 결과

4.1절의 경계공유 기준, 4.2절의 가까운 이웃 시군구 기준, 그리고 4.3절의 거리 기준 이웃 정보시스템에서 얻어진 모든 결과 중에서 대표적인 이웃정보시스템을 표 4.4에 정리하였다.

표 4.4를 살펴보면 경계를 공유하고 가중치를 모두 “1”로 정한 이웃정보시스템이 가장 큰 Moran's I를 주는 것을 확인 할 수 있다. 또한 경계를 공유하면서 가중치를 경계선의 길이에 비례하도록 정한 이웃정보시스템의 Moran's I 값이 다음으로 큰 것을 확인 할 수 있다. 가까운 이웃 시군구를 이웃으로 정하는 기준은 큰 Moran's I를 주지 못하고 있다. 특히 시군구 수를 4로 정하면 광주, 대전, 울산 광역시처럼 시군구 갯수가 5개인 지역에서는 모든 시군구가 이웃으로 정해져야 하는 문제가 발생한다. 거리 기준의 경우에는 특별시와 광역시, 경기도,

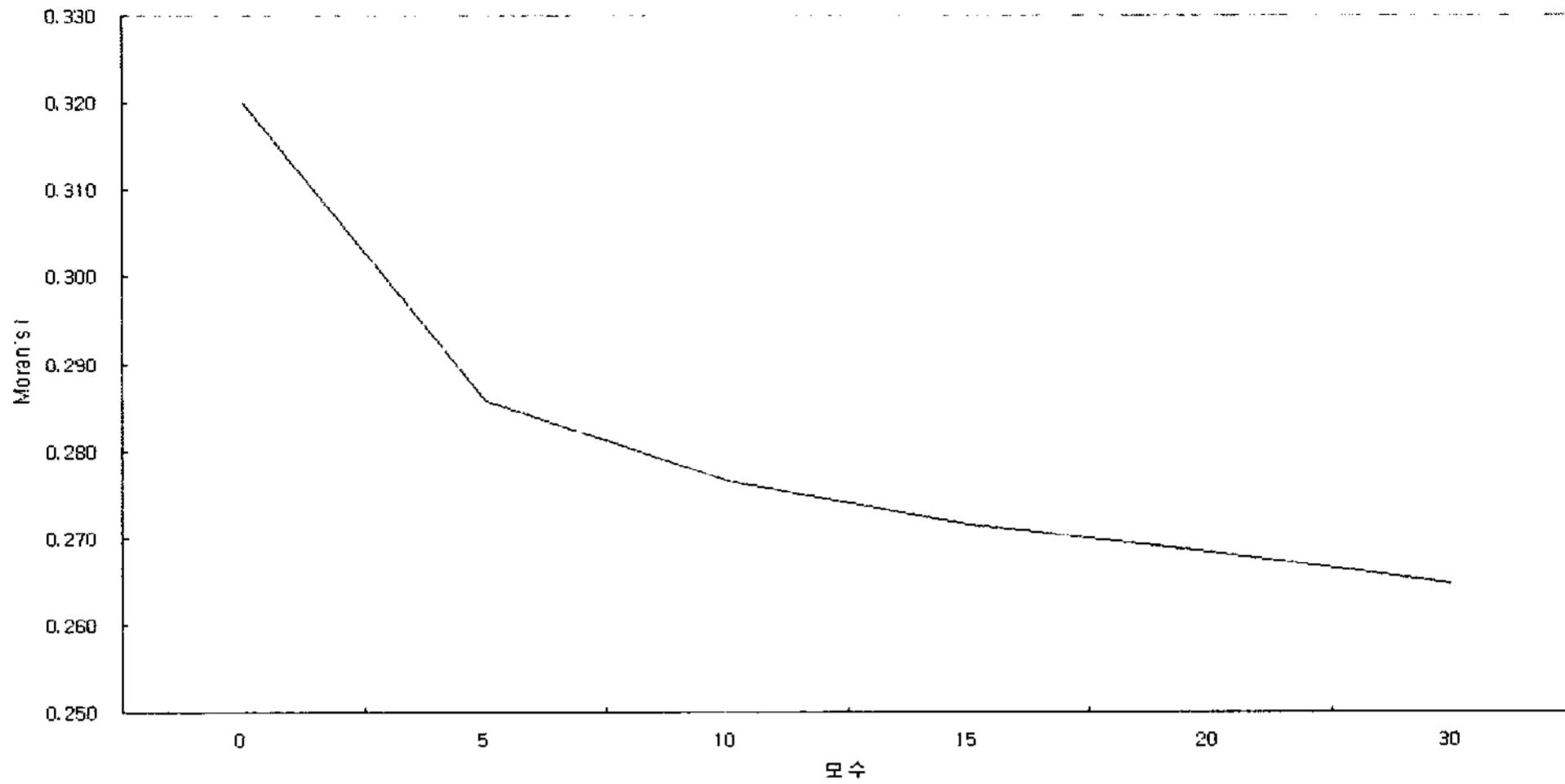


그림 4.4: $d_{ij}^{-\gamma}$ 의 모수 γ 에 따른 Moran's I 값 비교

표 4.4: 이웃정보시스템의 가중치별 Moran's I 비교

| 통계량 | 이웃 | 가중치 | 추정량 | 표준오차 |
|-----------|----------|--------------------------------|-------|-------|
| Moran's I | 경계 공유 | $w_{ij}=1$ | 0.372 | 0.051 |
| | | $w_{ij}=(l_{ik}/L_i)^{10}$ | 0.370 | 0.084 |
| | | $w_{ij}=(l_{ik}/L_i)/d_{ij}^2$ | 0.316 | 0.068 |
| | 시군 2개 이웃 | $w_{ij}=1$ | 0.310 | 0.061 |
| | 시군 3개 이웃 | $w_{ij}=1$ | 0.324 | 0.051 |
| | 시군 4개 이웃 | $w_{ij}=1$ | 0.282 | 0.045 |
| | 거리 기준 | $w_{ij}=1$ | 0.336 | 0.053 |
| | | $w_{ij}=1/k$ | 0.320 | 0.054 |

그외 도에 따라 거리 기준을 달리 정해야 하는 어려움이 있으면서도 Moran's I 값은 상대적으로 높지 않게 나왔다.

5. 결론

본 논문에서는 격자자료 분석에서 중요한 부분인 이웃정보시스템을 정하는 방법을 살펴보았다. 최근 격자자료 분석은 소지역 추정법 등에 응용되는 등 많은 분야에서 응용되고 있다. 이웃정보시스템을 정하기 위해서는 GIS 기법을 사용하는 것이 편리하나 가장 중요한 것은 자료에 관한 충분한 정보가 더욱 중요하다고 판단된다. 주어진 실업률 자료에 여러 이웃정보시스템을 적용한 결과 경계를 공유하면서 가중치를 모두 "1"로 정한 이웃정보시스템이 Moran's I 기준으로 가장 좋은 결과를 주었다. 그러나 다른 자료를 분석한다면 다른 결론이 얻어지게 되기 때문에 본 논문의 결과는 하나의 참고자료로 사용하는 것이 타당할 것이다.

참고문헌

- 김정오, 신기일 (2006). Comparison of small area estimations by sample sizes, *The Korean Communications in Statistics*, **13**, 669–683.
- 신기일, 이상은 (2003). Model-data based small area estimation, *The Korean Communications in Statistics*, **10**, 637–645.
- 황희진, 신기일 (2008). 축소 예측을 이용한 소지역 추정, <응용통계연구>, **21**, 109–123.
- Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*, Pion, London.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*, John Wiley & Sons, New York.
- Freeman, M. F. and Tukey, J. W. (1950). Transformation related to the angular and the square root, *The Annals of Mathematical Statistics*, **21**, 607–611.
- Kaluzny, S. P., Vega, S. C., Cardoso, T. P. and Shelly, A. A. (1998). *S+ Spatial Stats: User's Manual for Windows and UNIX*, Springer, New York.
- Rao, J. N. K. (2003). *Small Area Estimation*, John Wiley & Sons, New York.

[2008년 4월 접수, 2008년 4월 채택]

Comparison of Neighborhood Information Systems for Lattice Data Analysis*

Kang-Seok Lee¹⁾ Key-Il Shin²⁾

ABSTRACT

Recently many researches on data analysis using spatial statistics have been studied in various field and the studies on small area estimations using spatial statistics are in actively progress. In analysis of lattice data, defining the neighborhood information system is the most crucial procedure because it also determines the result of the analysis. However the used neighborhood information system is generally defined by sharing the common border lines of small areas. In this paper the other neighborhood information systems are introduced and those systems are compared with Moran's I statistic and for the comparisons, Economic Active Population Survey (2001) is used.

Keywords: Weight, neighborhood information, spatial correlation, conditional autoregressive spatial model.

* This research was supported by the research fund of Hankuk University of Foreign Studies, 2008.

1) Graduate student, Dept. of Statistics, Hankuk University of Foreign Studies, San89, Wangsan, Mohyun, Yongin, Kyonggi Do 449-791, Korea.

2) Corresponding Author. Professor, Dept. of Statistics, Hankuk University of Foreign Studies, San89, Wangsan, Mohyun, Yongin, Kyonggi Do 449-791, Korea.

E-mail: keyshin@hufs.ac.kr