

# The Spatial Correlation of Mode Choice Behavior based on Smart Card Transit Data in Seoul

Man Sik Park<sup>a,b</sup> · JinKi Eom<sup>c</sup> · Tae-Young Heo<sup>d,1</sup>

<sup>a</sup>Department of Statistics, Sungshin Women's University

<sup>b</sup>Institute of Statistics and Basic Science Research Institute, Sungshin Women's University

<sup>c</sup>Transport Systems Research Team, Korea Railroad Research Institute

<sup>d</sup>Department of Information and Statistics, Chungbuk National University

(Received April 18, 2013; Revised July 30, 2013; Accepted August 2, 2013)

---

## Abstract

In this study, we provide empirical evidence of whether a spatial correlation among mode choices at the TAZ(Traffic Analysis Zone) level exists based on transit smart card data observed in Seoul, Korea. The results show that the areas with a higher probability that passengers choose to take a bus are clustered and that those regions have fewer metro stations than bus stations. We also found that the spatial correlation turned out to be statistically meaningful and provided an opportunity for the potential use of the spatial correlation in modeling mode choices. A reliable spatial interaction would constitute valuable information for transportation agencies in terms of their route planning and scheduling based on the transit smart card data.

Keywords: Spatial association, spatial logistic regression model, mode choice, probability map.

---

## 1. 서론

자동화된 교통요금 징수시스템을 통해 얻어지는 교통카드 자료를 이용한 분석은 대중교통계획에 활발히 사용되고 있다. 교통카드자료는 기존의 표본조사를 통해 얻어진 자료들 보다 자료의 정확도 및 품질이 높고 대중교통을 이용하는 이용자에 대한 교통패턴에 대한 가치 있는 정보를 제공해 준다는 장점을 가지고 있어 요금정산, 노선계획 및 스케줄링 관점에서 보다 나은 대중교통 서비스를 제공하는데 필요한 다양한 연구가 필요하다.

기존 연구사례를 살펴보면 최근 몇 년간 교통카드자료를 이용한 연구들이 활발히 진행되고 있다 (Bagchi와 White, 2005; Utsunomiya 등, 2006; Bryan과 Blythe, 2007; Chu와 Chapleau, 2008; Eom 등, 2009; Jang, 2010). Blythe (2004)는 교통카드자료를 기반으로 통행수요가 대중교통 네트워크상에 어떻게 발생하고 있는지를 분석하여 대중교통 이용패턴을 설명하였으며, Bagchi와 White (2005)는 대중교통 운영자측면에서 대중교통 이용자 그룹별로 통행패턴을 분석하고 대중교통 이용활성화를 위한 방안에 대한 연구를 수행하였다. Chu와 Chapleau (2008)은 시공간 이용패턴을 구현하여 지

---

<sup>1</sup>Corresponding author: Associate Professor, Department of Information and Statistics, Chungbuk National University, Cheong-Ju, Chungbuk 361-763, Korea, E-mail: theo@cbnu.ac.kr

**Table 2.1.** Characteristics of Transit Smart Card Data

DATE	# of Card data	# of Error data	% of Error	# of Data used
2009/10/25	11,113,808	470,218	4.23%	10,643,590
2009/10/26	18,440,177	758,304	4.11%	17,681,873
2009/10/27	18,545,592	755,997	4.08%	17,789,595
2009/10/28	18,560,858	753,533	4.06%	17,807,325
2009/10/29	18,346,370	742,397	4.05%	17,603,973
2009/10/30	18,611,267	756,965	4.07%	17,854,302
2009/10/31	13,755,745	603,199	4.39%	13,152,546

역별 탑승시간에 따른 대중교통 네트워크 이용정도를 제시하여 대중교통 서비스 향상을 위해 노선조정 및 운행스케줄을 합리적으로 조정할 필요성을 제기하였다. 국내 연구사례로서 Jang (2010)은 교통카드 자료를 이용하여 대중교통 이용자의 통행시간 및 환승패턴을 다양한 GIS 지도를 통해 표출하여 대중교통 환승 및 서비스 개선에 활용방안을 제시하였다. Lim 등 (2012)은 교통카드자료를 이용하여 통행자의 지역별 통행분포를 추정하는 연구를 수행하였다. Eom 등 (2012)은 도시철도의 정시성과 차내혼잡에 대한 서비스 분석을 수행하였고, Eom 등 (2013)은 명절기간 서울역 철도 이용객의 접근통행 특성을 서울시 25개 구단위로 비교 분석 하였다.

교통카드자료를 이용한 다양한 연구들이 수행되었으나 현재까지 대중교통 이용자들이 대중교통 수단선택과 지역적인 공간적 특성에 관한 연구는 미흡한 실정이다. 따라서 본 연구에서는 대중교통 통행자들이 이용하는 대중교통 수단과 출발지에 대한 공간적 상관성 여부를 실증적으로 분석하여 향후 대중교통 수단선택 모형에 이를 반영하는 것에 대하여 검토하도록 한다. 대중교통수단선택에 있어 이용자의 출발지에 대한 공간상관성 분석결과는 향후 대중교통 노선계획 및 서비스 확충에 있어 큰 도움이 될 것으로 판단된다.

## 2. 교통카드자료

서울시는 2004년에 대중교통 요금징수 자동화 시스템을 도입하여 현재까지 대중교통요금을 이용자 거리 및 환승에 따라 다양한 요금을 징수하고 있다. 교통카드자료에 의하면 2009년 현재 대중교통 이용자의 95%가 교통카드를 이용하는 것으로 분석되었으며 최대 4번의 환승할인에 대한 대중교통요금정책에 따른 결과로 이해된다.

본 연구에 사용된 교통카드자료는 서울시 통합교통요금에 대한 대중교통 운영주체별 요금정산 대행 기관인 한국스마트카드사로부터 2009년 10월 25일부터 31일까지 일주일 자료를 확보하여 분석 하였다. Table 2.1은 일주일 자료를 분석한 내용으로 일 평균 자동징수시스템에 기록된 자료는 약 3천6백만건에서 3천8백만건인 것으로 나타났다. 각각 기록된 건수는 승차 및 하차 시에 대중교통카드를 요금징수 패드에 터치할 경우 발생된 것으로 출발점과 도착점에 각각 기록된 통행량으로 환산할 경우 약 천8백만에서 천9백만 통행에 이르게 된다. 자료분석시 발견된 오류는 승하차 태그오류, 승하차 동일역/정류장 태그, 통행시간 오류 등으로 일 평균 전체 관측자료의 약 4%정도를 차지하는 것으로 분석되었으며 본 연구에서 제외하였다. 따라서 최종 연구에 활용된 자료는 약 천7백만 통행으로서 이는 서울시의 약 2만 개 이상의 버스정류장과 300개 도시철도역에서 기록된 것이다. 본 연구에서는 모든 자료에 대해 공간상관성 분석을 수행하기에 컴퓨터 연산이 현실적으로 어려우므로 이들 개별 자료를 GIS를 이용하여 서울시 465개 행정동 단위로 집합화하였다 (Goodchild, 1987; Ding과 Fotheringham, 1992; Goodchild 등, 1992).



Figure 2.1. Metro Lines in Seoul (revised from original map on <http://naver.com>)

Figure 2.1은 서울시 도시철도 노선도를 보여 주고 있다. 그림에서 보는바와 같이 도시철도 서비스가 제공되지 않는 지역에서는 버스의 이용확률이 높을 것으로 예상된다. 서울시의 경우 서울시 강북지역은 대부분 산지로 되어있고 인구밀도 낮은 관계로 도시철도 서비스가 제한적임에 따라 서울시 남부에 비해 버스의 이용도가 높을 것으로 판단된다.

### 3. 모형

#### 3.1. 공간 로지스틱 회귀모형(spatial logistic regression model)

이분형(binary) 자료에 대한 로지스틱 회귀모형(logistic regression model)은 일반화선형모형(generalized linear model)에서 로짓(logit)이라는 연결함수(link function)를 통한 확률변수로 간주될 수 있다. 지역  $i$ 에서 관측된  $m_i$ 개의 관측값에 대한 성공의 횟수를  $Y_i$ 라고 하면 본 연구에서의  $Y_i$ 는 지역  $i$ 에서 버스를 선택하는 횟수(승객 수)로 정의할 수 있다. 모든  $i = 1, 2, \dots, n$ 에 대하여

$$Y_i \sim B(m_i, \pi_i),$$

$$\text{logit}(\pi_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \mathbf{x}_i^T \boldsymbol{\beta} + \delta_i$$

여기서  $\pi_i$ 는 지역  $i$ 에서 버스를 선택할 확률을 나타내며,  $\boldsymbol{\beta}$ 는  $p$ 개의 설명변수  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ 에 대응하는 모수벡터이며,  $\delta_i$ 는 모든  $i$ 에 대해 평균이 0인 확률과정으로 정의할 수 있다. 본 연구에서는 기본적인 이항회귀모형에 공간과정(spatial process)인  $\delta_i$ 를 반영하여 공간상관성을 확인할 수 있는 모형을 제안하고자 한다.

### 3.2. 조건부 자기회귀(conditional autoregressive; CAR)모형

공간 로지스틱 회귀모형에 포함되는 공간과정인  $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)^T$ 는  $n$ 차원의 다변량 정규분포(multi-variate normal distribution)를 가정하는 다음과 같은 조건부 자기회귀모형으로 정의된다.

$$\boldsymbol{\delta} \sim N_n(\mathbf{0}, \boldsymbol{\Omega}) \quad (3.1)$$

여기서  $\mathbf{0} = (0, \dots, 0)^T$ 는 영벡터이고  $\boldsymbol{\Omega} = \{w_{ij}, i, j = 1, \dots, n\}$ 는 공분산행렬(covariance matrix)을 의미한다. 공간상관성을 반영한 공간과정은 조건부분포(conditional distribution)의 형태로 일반화할 수 있으며, 이를 조건부 자기회귀모형이라고 한다.

조건부자기회귀모형의 장점은 첫째, 지역 간 공간상관성에 대한 효과를 측정할 때 인접정보 또는 거리 정보 등을 이용하여 이웃 정보에 대해 쉽게 접근할 수 있으며 둘째, 조건부자기회귀모형이 전통적인 모수 추정방법인 최우추정(maximum likelihood estimation)방법과 베이시안(Bayesian) 추정방법에 모두 적용될 수 있다. Best 등 (1999)과 Conlon과 Waller (2000)는 거리기반 이웃정보를 활용하여 베이시안 기법을 기초로 한 조건부 자기회귀모형을 사용하였다. 또한 Cressie 등 (1999)과 Kaiser 등 (2002)은 지구통계 자료에 조건부 자기회귀모형을 이용하였으며, 대기오염 정도의 공간적 예측에 활용하였다.

### 3.3. 공간과정 및 이웃구조

식 (3.1)의 조건부 자기회귀모형을 구현함에 있어서 지역들 간의 연관성의 정도를 공분산행렬로 표현하기 위해 이웃행렬(neighborhood matrix)에 대한 정의가 필요하다. 이 논문에서는 두 가지의 이웃행렬에 대해 정의하고 분석에 활용하고자 한다. 각 이웃행렬에 대해 공간과정에 대한 확률분포는 다음의 식 (3.2)와 같이 정의할 수 있다.

$$\delta_i | \{\delta_j, \tau^2\} \sim N \left( \rho \sum_{i \sim j} c_{ij} \delta_j, \frac{\tau^2}{k_{ii}} \right) \quad (3.2)$$

여기서  $i \sim j$ 는 지역  $j$ 가 지역  $i$ 의 이웃에 속하는 것을 나타내며,  $k_{ii}$ 는 지역  $i$ 에 이웃하는 지역의 개수를 의미한다. 모수  $\rho$ 와  $\tau^2$ 은 공간상관계수(spatial correlation coefficient)과 조건부 분산(conditional variance)을 각각 의미한다. 식 (3.2)는 인수분해 정리(factorization theorem)에 의하여 식 (3.3)과 같이 결합확률분포(joint probability distribution)로 나타낼 수 있다.

$$\boldsymbol{\delta} \sim N(\mathbf{0}, \tau^2 [\mathbf{D}(1 - \rho\mathbf{C})]^{-1}) \quad (3.3)$$

여기서  $\mathbf{D}$ 는 이웃하는 지역의 개수( $k_{ii}, i = 1, \dots, n$ )를 대각원소로 하는  $n \times n$  대각행렬을 의미한다. 이웃관계 여부를 규정짓는 행렬  $\mathbf{W} = \{w_{ij}\}$ 에 대하여 임의의 지역  $i$ 의 이웃인  $j$ 는  $w_{ij} = 1$ 이고 그렇지 않은 경우  $w_{ij} = 0$ 을 가진다. 이로부터 행렬  $\mathbf{C} = \mathbf{D}^{-1}\mathbf{W}$ 는 이웃관계를 가지는 특정한  $(i, j)$ 에 대해 지역  $i$ 의 전체 이웃의 개수로 나누어준, 조정된 인접행렬(adjacency matrix)을 의미한다.

본 연구에서는 서울시 행정동 자료를 이용하여 공간수단선택모형을 개발하기 위해 공간과정으로 조건부 자기회귀모형(conditional autoregressive model)을 이용하였다. 이 논문에서 사용하는 이웃행렬은 행정동의 경계면을 공유하는 것으로 정의하는 구조(first-order neighborhood structure)와 지역의 중심점 간의 거리가 2.3km 이내로 정의하는 구조(distance-based neighborhood structure)를 각각 고려한다. 여기서, 2.3km는 모든 지역이 적어도 하나의 이웃을 가지는 최소한의 거리이다. 즉, 첫 번째 이웃행렬에서는 서울시 각 행정동 별로 경계선을 인접한 이웃으로 하나의 행정동에 경계선이 인접한 모든 행정동은 이웃으로 간주하는 구조이다. 두 번째 이웃행렬에서는 거리 기반의 이웃행렬로서 하나의 행정동의

중심점에서 적어도 하나 이상의 이웃을 가지는 최소거리로 설정한 2.3km 이내의 모든 행정동들은 이웃으로 간주하는 구조이다.

### 3.4. 계층모형구조(hierarchical modeling structure)

이분형 반응변수에 대한 공간 로지스틱 회귀모형과 공간과정에 부여하는 조건부 자기회귀모형을 결합하기 위해 다음과 같은 계층모형구조를 제안하고자 한다. 다음과 같은 계층모형구조의 모수 추정을 위해 베이저안 기법을 이용한다.

$$\begin{aligned} Y_i | \pi_i &\sim B(m_i, \pi_i) \\ \pi_i | \boldsymbol{\beta}, \delta_i &\sim \text{logistic}(\boldsymbol{\beta} + \delta_i) \\ \boldsymbol{\beta} &\sim N(0, \sigma^2) \\ \boldsymbol{\delta} | \tau^2, \rho &\sim N(0, \tau^2 [\mathbf{D}(\mathbf{I} - \rho \mathbf{C})]^{-1}) \\ \tau^2 &\sim \text{IGamma}(a, b) \\ \rho &\sim \text{Uniform}\left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}}\right) \end{aligned}$$

여기서, IGamma는 역감마분포(Inversed gamma distribution),  $\lambda_{(1)}, \lambda_{(n)}$ 은 각각 공간인접행렬의 최소 및 최대 고유값을 의미한다. 위 식에서 보는 것과 같이 이동수단으로 버스를 선택할  $i$ 번째 지역의 승객 수,  $Y_i$ 는 이항분포를 따르는 것으로 가정할 수 있고 이 때 성공의 확률은 버스를 선택할 확률이 되겠다. 따라서 일반화선형모형의 이론적 배경에 의해 성공의 확률에 영향을 주는 체계적 성분(systematic component)은 여러 잠재변인들의 선형결합과 공간상관성을 띄는 오차항의 합으로 이루어지고 랜덤성분(random component)와의 연결함수는 로짓연결(logit link)를 사용한다. 이 연구에서는 체계적 성분으로 지역의 위치정보에 상관없이 일정한 모수값을 갖는 형태로 단순화하여 구성하였다. 체계적 성분의 다른 한 축을 차지하는 오차항,  $\{\delta_i\}$ 는 조건부 자기회귀모형을 따르게 되고 이때 이웃구조는 앞에서 언급한 바와 같이 두 가지의 이웃행렬을 고려한다. 조건부 자기회귀모형에서의 조건부 분산은 역감마분포를 따르고 공간상관계수인  $\rho$ 는 오차항  $\boldsymbol{\delta}$ 의 공분산행렬의 고유값들로 이루어진 범위 하에서의 균일분포를 따른다고 가정한다 (Banerjee 등, 2004).

## 4. 자료분석

계층모형구조를 교통카드자료에 적용하기 위하여 우선 공간상관성 정도를 평가하고자 한다. 이를 바탕으로 절편항만을 포함한 계층적 모형을 고려하고자 한다.

### 4.1. 탐색적 자료분석

지역들 간의 상관성을 내포하는 자료에 공간 로지스틱 회귀모형을 적용하기에 앞서 먼저 공간상관성을 측정 및 확인해야 할 필요가 있겠다. 일반적으로 상관계수는 두 변수 사이에 존재하는 상관성의 정도를 나타내지만 공간자기상관계수는 두 변수간의 상관성이 아닌 하나의 변수에 공간상의 여러 개의 관측값들 사이에 존재하는 연관성을 나타낸다는 점에서 일반적인 상관계수와는 다른 특성을 가지고 있다. 본 연구에서 공간 상관성을 나타내는 측도로 사용할 공간자기상관계수에 대하여 간략히 소개하면 다음과 같다. 공간자기상관계수는 어떤 공간 현상이 이웃 지역과 유사한 정도를 정량적으로 표현하는 지수를 의미한다 (Goodchild, 1986). 공간상에 분포해 있는 자료들 간의 이웃정보를 정의하는 방법에는 주로

**Table 4.1.** Moran's I and Geary's C for Spatial Dependence

	Estimate	Std. error	P-value
Moran's I	0.0617	0.02872	0.0263
Geary's C	0.9392	0.02910	0.0368

거리척도와 인접성 척도를 기준으로 하고 있으며, 본 연구에서는 인접성 척도를 사용하였다. 특히 이러한 공간자기상관계수를 얻기 위해서는 단위 지역들 사이의 공간적 이웃관계에 대해 정의하는 이웃정보 행렬의 구성이 선행되어야 하며, 이는 단위지역의 개수가 행렬의 크기인 정방행렬,  $\mathbf{W}$ 로서 단위지역  $i$ 와  $j$ 간의 이웃 정보를 나타내는 행렬이다 (Besag, 1974; Sokal과 Oden, 1978).

대표적인 공간연관성 지수로는 공분산을 기반으로 하는 Moran의 Index (Moran, 1950; Cliff와 Ord, 1972, 1981)가 있으며 아래와 같이 정의된다.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\left( \sum_{i \neq j} w_{ij} \right) \sum_{i=1}^n (z_i - \bar{z})^2}$$

여기서,  $n$ 은 단위 지역의 수,  $z_i$ 와  $z_j$ 는 각각  $i$ 지역과  $j$ 지역에서의 버스를 이용한 비율(proportion)을 나타내며,  $\bar{z}$ 는 전체 지역에서의 버스이용비율의 산술평균이다.  $w_{ij}$ 는 공간가중치행렬  $\mathbf{W}$ 의  $i$ 지역과  $j$ 지역 사이의 공간 가중치 원소를 의미한다.

Moran의 Index는 공간적 자기 상관을 탐색하는데 유용한 측정지표로서 인접한 공간자료들을 비교하여 통계량을 계산하며 일반적인 상관관계와 마찬가지로  $-1$ 과  $1$ 사이의 값을 가진다. Moran의 Index는 공간상관이 양의 상관이 강할수록  $1$ 에 가까우며 음의 상관이 강할수록  $-1$ 에 가깝고  $0$ 은 공간상관성이 존재하지 않음을 의미한다.

Moran Index와 비슷하게 Geary의  $C$  통계량을 통해서도 공간상관성을 파악할 수 있으며, Geary의  $C$  통계량은  $0$ 과  $1$  사이의 값을 갖는다 (Geary, 1954). Geary의  $C$  통계량이  $1$ 로 근접할수록 양의 상관관계를 가지고 된다. 통계량의 형태는 다음과 같이 표현된다.

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2}{\left( \sum_{i \neq j} w_{ij} \right) \sum_{i=1}^n (z_i - \bar{z})^2}$$

공간분석을 수행하기 전에 지역간 공간적 상관성의 존재 여부에 대한 통계적 검정을 위해 Moran의 Index와 Geary의  $C$  통계량을 통해 확인하였다. Table 4.1를 통해 알 수 있듯이 지역 간 버스를 선택할 비율에 대해서 공간적으로 상관되어 있음을 알 수 있다. 공간과정을 모형에 반영한 공간 로지스틱 회귀모형은 신뢰성 있는 수단분담률 추정을 위해 필요할 것으로 판단된다.

#### 4.2. 계층모형의 적합결과

본 연구에서 제시한 이웃행렬의 형태로 구분된 두 개의 모형의 모수를 추정하기 위하여 마코프 연쇄 몬테칼로 방법을 사용하여 범용 베이저안 통계 패키지인 WinBUGS (Spiegelhalter 등, 2000)를 사용하였으며, 추정된 모수 값의 수렴상태를 알아보기 위하여 세 가지의 서로 다른 초기치를 사용하였으며, 이를 통한 사후분포 표본을 통하여 모수 추정값을 얻을 수 있었다. 추정된 모수의 올바른 수렴 여부를 확인하

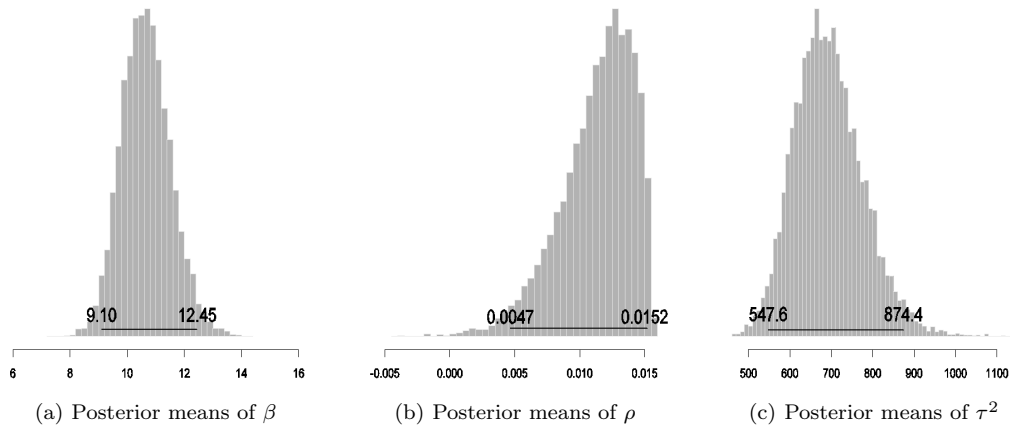


Figure 4.1. Histograms of the Posterior Means of the Parameters and the 95% Bayesian Confidence Intervals (solid lines) based on a Boundary-based Neighborhood.

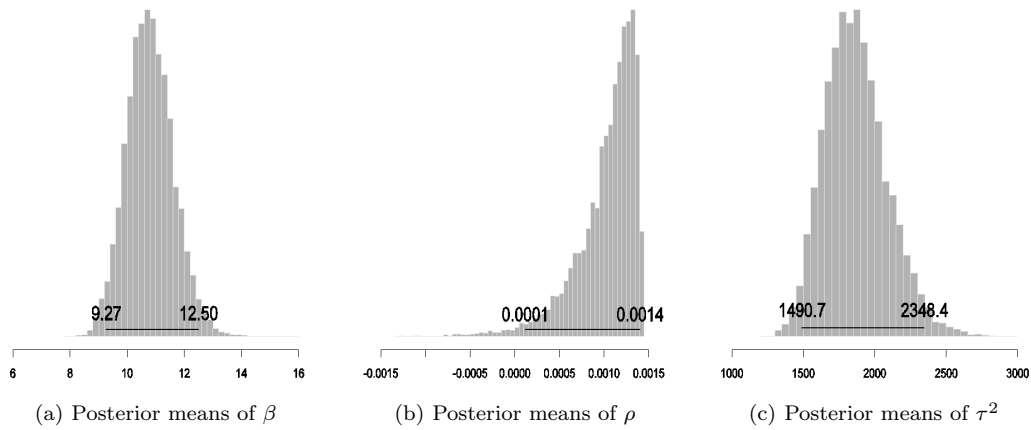
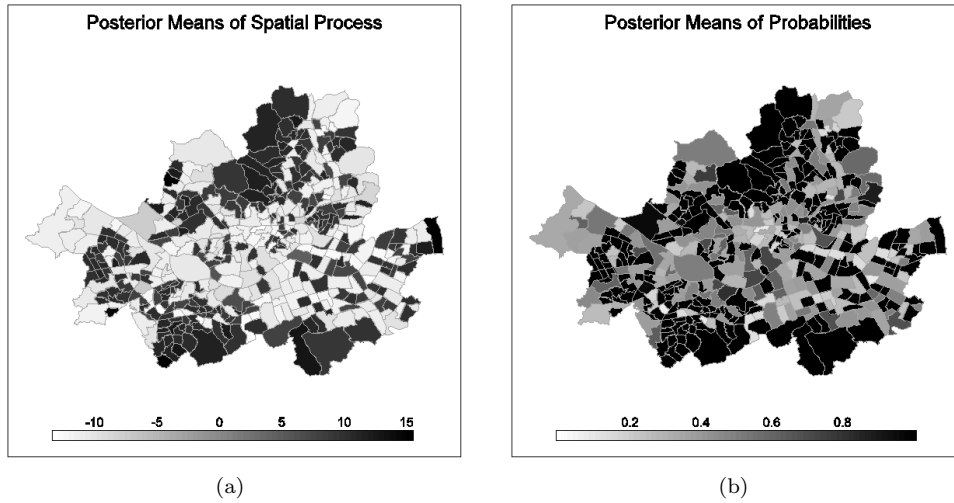


Figure 4.2. Histograms of the Posterior Means of the Parameters and the 95% Bayesian Confidence Intervals (solid lines) based on Distance-based Neighborhood.

기 위해 Gelman과 Rubin (1992) 방법을 이용하여 모수에 대한 수렴진단을 하였다. 본 연구에서는 제시하지 않았지만 공간로지스틱모형에 대하여 세 개의 초기값들을 적용한 결과 모수들의 사후분포 표본들에 대한 Gelman-Rubin 수렴진단 측도가 1에 가깝게 나타나고 있으므로 생성된 모수들이 잘 수렴하여 올바른 추정결과임을 알 수 있다. 모수추정의 결과는 세 개의 초기값을 활용하여 각각 20000번 반복한 후에 각 3000번까지 제거한 나머지 51000번의 결과값을 사후분포의 표본으로 이용하였다.

Figure 4.1 and 4.2는 이웃구조에 따른 모수들의 사후분포 하에서의 추정값에 대한 히스토그램과 신뢰구간에 대한 결과이다. 여기서 주목할 만한 결과는 이웃구조를 어떻게 정의하느냐에 상관없이 공간상관성은 통계적으로 유의하고 양(+)의 값을 가지는 것으로 나타났다. 다시 말해서, 특정지역에서 출발하는 승객들 중 지하철보다 버스를 보다 많이 이용할수록 이웃한 지역 역시 버스를 선택하는 승객이 더 많게 됨을 의미한다. 조건부 분산의 경우 이웃구조를 어떻게 정의하느냐에 따라 특정한 지역의 이웃의 개수가 달라지므로 추정값과 신뢰구간의 값에 다소 차이가 있다. 사후분포의 결과를 토대로 각 지역에서의



**Figure 4.3.** Posterior Means of Spatial Process (a) and Probabilities of Choosing a Bus (b) based on Boundary-based Neighborhood structure (Data observed at 2009/10/28).



**Figure 4.4.** Posterior Means of Spatial Process (a) and Probabilities of Choosing a Bus (b) based on Distance-based neighborhood structure (Data observed at 2009/10/28).

$\delta_i$ 와 성공의 확률인  $\pi_i$ 의 사후평균값들을 이용하여 각 이웃구조에 따른 예측지도를 구성하였다 (Figure 4.3, 4.4).

그림을 통해 알 수 있듯이 이웃구조의 형태에 상관없이, 사후추정값이 높은 지역들이 군집을 형성하고 있음을 알 수 있다. 이는 양의 공간상관성이 유의하게 나타났음을 증명하는 것으로 크게 3개의 군집이 존재하는 것으로 나타났다(서울 북부, 2군데의 서울남부). 이 영역들에서는 공간상관성을 내포하는 오차항의 추정값과 버스를 선택할 확률의 추정값 모두 타 지역에 비해 높게 나타났다. 현재 서울시의 지하철선도와 비교해 본다면 이 현상은 쉽게 이해할 수 있다. 그 이유는 이 영역들은 주로 산악지역으로 이루어져 있어서 지하철보다는 마을버스나 노선버스의 접근이 상대적으로 용이한 상태여서 승객들이 지



하철을 이용할 수 있는 환경이 조성되어 있지 않다. 이러한 지역의 특수성이 오차함에 의해 설명되어 지므로 이를 고려하지 않고 전통적인 회귀분석을 사용하게 될 경우 자료의 설명력이나 예측력이 떨어지게 됨을 명확히 알 수 있다.

일반적으로 지하철역이 도보로 쉽게 접근할 수 있는 거리 이내에 있다면 대중교통을 이용하는 사람들은 지하철을 선택하는 경향이 있고 그렇지 않은 경우에는 버스를 선호하게 된다. 서울은 우리나라에서 교통 공학의 관점에서 가장 발전된 도시이고 특히 대부분의 주요 지점들이 지하철로 잘 연결되어 있다. 하지만 이러한 주요 지점에서조차도 2~3개의 지하철역만이 존재하고, 특히 평균적으로 30개 이상의 버스정류장과 비교하면 턱없이 적은 실정이다. 이는 지하철노선의 신설 및 관리에 있어서 막대한 예산이 투입되는 데 반해 버스노선의 경우 상대적으로 작은 비용으로 인구밀도가 낮거나 접근성이 용이하지 않은 지역에서도 운영할 수 있다는 점에 기인한다고 할 수 있다. 하지만 지하철에 의한 교통수단 부담을 은 대량운송 수단이고 운행시간이 정확하며 이동시간이 비교적 짧다는 점으로 인해 버스와 거의 동일한 수단부담을 가진다고 볼 수 있다. 종합적으로 지하철과 버스노선이 상호 보완적으로 기능을 수행하기 때문에 서울시의 대중교통 수단부담을 향상시키고 있음을 알 수 있다.

## 5. 결론

본 연구에서는 서울시 대중교통카드 자료를 이용하여 승객들의 대중교통 수단선택에 있어 공간적 상관성 존재 여부를 검증하였으며 공간상관성을 반영한 수단부담 모형 구축의 가능성을 검토하였다. 철도와 버스의 두 가지 수단을 중심으로 철도역과 버스정류장의 위치정보를 활용하여 공간상관성을 존재 여부를 검증하기 위해 두 가지 이웃구조인 유클리디언 중심점 거리구조와 경계면 공유 구조에 기반한 공간가중치를 적용하여 테스트 하였다. 두 가지 이웃구조 자료는 서울시 행정동별로 철도와 버스 정류장을 집합화하여 구성하였으며 통계적 분석결과 대중교통 수단선택에 있어 공간상관성의 존재를 확인 하였다.

서울시의 동별로 철도와 버스의 선택 확률을 분석한 결과 버스선택 확률이 높은 지역이 서로 공간적 그룹을 형성하는 것으로 나타났으며 이러한 지역의 공통된 특성은 지하철 역의 수 보다는 버스 정류장의 수가 상당히 많은 것으로 나타났다. 이러한 패턴은 공간상관성을 반영한 수단선택 예측 그림에서도 양호하게 반영됨에 따라 향후 대중교통카드 자료를 기반으로 공간상관성을 반영한 수단선택 모형의 구축이 가능할 것으로 판단된다. 기존 수단선택 모형은 설문조사 방식으로 획득된 자료에 기반하여 통행시간과 비용의 함수로 정의 되어 승객의 탑승 위치에 상관없이 동일한 통행시간과 비용이 소요될 경우 같은 대중교통 확률을 제시하는 한계를 가지고 있다. 따라서 개별 승객의 출발지를 기반으로 한 공간상관성 효과의 대중교통 수단선택 모형의 반영은 실제 승객의 수단선택 현실을 보다 정확하게 반영할 수 있을 것으로 예상된다.

결론적으로 대중교통카드 자료의 활용은 지리적으로 이용객의 철도 및 버스의 탑승위치를 제공하는 유용한 자료로 판단되며 이를 통해 대중교통 수단선택에 있어 공간적 영향력을 계량화 하여 반영할 수 있으며 공간상관성을 반영한 대중교통 수단선택 모형은 보다 정확하고 현실성이 있는 수단선택 확률을 예측하는데 상당한 도움을 줄 것으로 예상된다. 향후 공간상관성을 반영한 수단선택모형을 구축하여 기존 모형구축 방식에 비해 예측력이 우수함을 검증하는 연구가 필요하며 개선된 모형은 대중교통계획 및 서비스를 위해 중요한 수단으로 활용될 것으로 기대된다.

## References

- Bagchi, M. and White, P. R. (2005). The potential of Public Transport Smart Card Data, *Transport Policy*, 12, 464-474.

- Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton, Florida.
- Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society B*, **36**, 192–225.
- Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A. and Conlon, E. M. (1999). Bayesian models for spatially correlated disease and exposure data, *Bayesian Statistics*, Oxford University Press, Oxford, **6**, 131–156.
- Blythe, P. (2004). Improving Public Transport Ticketing through Smart Cards, *Proceedings of the Institute of Civil Engineers: Municipal Engineer*, **157**, 47–54.
- Bryan, H. and Blythe, P. (2007). Understanding behavior through Smart card Data Analysis, *Proceedings of the Institute of Civil Engineers: Transport*, **160**, 173–177.
- Chu, K. K. and Chapleau, R. (2008). Enriching Archived Smart Card Transaction Data for Transit Demand Modeling, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2063, Transportation Research Board of the National Academies, Washington, D.C., 63–72.
- Cliff, A. and Ord, J. K. (1972). Testing for Spatial Autocorrelation among Regression Residuals, *Geographical Analysis*, **4**, 267–284.
- Cliff, A. and Ord, J. K. (1981). *Spatial Processes, Models and Applications*, Pion Limited, Las Vegas.
- Conlon, E. M. and Waller, L. A. (2000). Flexible spatial hierarchical models for mapping disease rates”, *American Statistical Association Proceedings of the Section on Statistics and the Environment*, 82–87.
- Cressie, N., Kaiser, M. S., Daniels, M. J., Aldworth, J., Lee, J., Lahiri, S. N. and Cox, L. H. (1999). Spatial analysis of particulate matter in an urban environment, in Gomez-Hernandez, J., Soares, A. and Froidevaux, R. (Eds.): *geoENV II - Geostatistics for Environmental Applications*, Kluwer, Dordrecht, 41–52.
- Ding, Y. and Fotheringham, A. S. (1992). The integration of spatial analysis and GIS, *Computers, Environment and Urban Systems*, **16**, 3–19.
- Eom, J. K., Choi, M. H., Kim, D. S., Lee, J. and Song, J. Y. (2009). Evaluation of Metro Services based on Transit Smart Card Data (A Case Study of Incheon Line 1), *Journal of The Korean Society for Railway*, **15**, 80–87.
- Eom, J. K., Lee, J. and Lee, K. S. (2013). Access and Egress Patterns of Travel to a Regional Railway Station Based on Transit Smart Card Data (Case study: Seoul Station during Chuseok Holiday), *Journal of The Korean Society for Railway*, **16**, 59–64.
- Eom, J. K., Park, M. S. and Heo, T. Y. (2012). Estimating Probability of Mode Choice at Regional Level by Considering Spatial Association of Departure Place, *Journal of The Korean Society for Railway*, **12**, 656–662.
- Geary, R. C. (1954). The contiguity ratio and statistical mapping, *The Incorporated Statistician*, **5**, 115–145.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences, *Statistical Science*, **7**, 457–511.
- Goodchild, M. F. (1986). *Spatial Autocorrelation*, CATMOG 47, Geobooks: Norwich UK.
- Goodchild, M. F. (1987). A spatial analytical perspective on geographical information systems, *International Journal of Geographical Information Systems*, **1**, 335–354.
- Goodchild, M. F., Haining, R. P. and Wise, S. M. (1992). Integrating GIS and spatial data analysis: Problems and possibilities, *International Journal of Geographical Information Systems*, **6**, 407–423.
- Jang, W. (2010). *Travel Time and Transfer Analysis Using Transit Smart Card Data*, CD-ROM. Proceedings of 89th Annual Transportation Research Board Meeting, Washington, D.C.
- Kaiser, M. S., Daniels, M. J., Furukawa, K. and Dixon, P. (2002). Analysis of particulate matter air pollution using Markov random field models of spatial dependence, *Environmetrics*, **13**, 615–628.
- Lim, Y. T., Park, C., Kim, D. S., Eom, J. K. and Lee, J. (2012). Estimating Trip Distribution Model by Using Transit Card Data, *Journal of Transport Research*, **19**, 1–11.
- Moran, P. (1950). Notes on continuous stochastic phenomena, *Biometrika*, **37**, 17–23.
- R Development Core Team. (2005). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, (Available from [www.R-project.org](http://www.R-project.org)).
- Sokal, R. R. and Oden, N. L. (1978). Spatial autocorrelation in biology, I. Methodology, *Biological Journal of the Linnæan Society*, **10**, 199–228.

- Spiegelhalter, D. J., Thomas, A. and Best, N. G. (2000). WinBUGS Version 1.4 User Manual, Cambridge: Medical Research Council Biostatistics Unit, 2000. (Available from <http://www.mrc-bsu.cam.ac.uk/bugs>)
- Utsunomiya, M., Attanucci, J. and Wilson, N. H. (2006). Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, Transportation Research Record: Journal of the Transportation Research Board, No. 1971, Transportation Research Board of the National Academies, Washington, D.C., 119–126.

# 교통카드 자료를 이용한 서울시 지역별 대중교통 수단 선택 공간상관성 분석

박만식<sup>a,b</sup> · 엄진기<sup>c</sup> · 허태영<sup>d,1</sup>

<sup>a</sup>성신여자대학교 통계학과, <sup>b</sup>성신여자대학교 통계연구소 및 기초과학연구소  
<sup>c</sup>한국철도기술연구원 교통체계분석연구단, <sup>d</sup>충북대학교 정보통계학과

(2013년 4월 18일 접수, 2013년 7월 30일 수정, 2013년 8월 2일 채택)

---

## 요약

본 연구에서는 교통 분석존(서울시 행정동) 단위별로 대중교통 수단(버스, 도시철도)선택에 있어서 공간 상관성이 존재하는지 여부를 대중교통카드 자료를 기반으로 제시한다. 분석결과 버스를 탑승한 비율이 높은 지역들이 서로 이웃하여 그룹을 형성하고 있으며, 이들 지역은 도시철도 역사의 수가 버스 정류장에 비해 매우 적기 때문인 것으로 분석되었다. 버스에 탑승한 비율이 비슷한 그룹 간에는 공간 상관성이 존재하는 것으로 통계분석결과 나타났으며, 이러한 공간상관성은 향후 대중교통 수단선택 모형 구축에 고려할 수 있을 것으로 판단된다. 대중교통 수단선택에 있어 공간상관성의 존재는 대중교통 운영기관이 향후 대중교통카드를 기반으로 대중교통 노선계획, 운영계획을 수립함에 있어 중요한 정보가 될 것으로 기대된다.

주요용어: 공간연관성, 공간로지스틱회귀모형, 수단선택, 확률지도.

---

<sup>1</sup>교신저자: (361-763) 충북 청주시 흥덕구 개신동, 충북대학교 정보통계학과, 교수. E-mail: theo@cbnu.ac.kr