

공간로지스틱 모형을 이용한 이용자 출발지 기준 수단선택 확률추정 연구

Estimating Departure-based Mode Choice by Spatial Logistic Models

엄진기†
Jin Ki Eom,

문대섭*
Dae Seop Moon,

박만식**
Man Sik Park,

허태영***
Tae-Young Heo

Abstract

In general, the analysis of travelers' mode choice behavior is accomplished by developing the utility functions which reflect individual's preference of mode choice according to their demographic and travel characteristics. In this paper, we propose a methodology that takes the spatial effects of individuals' departure locations into account in the mode choice model. The proposed methodology allows us to estimate mode shares by departure places even though the survey does not cover all areas. This will help transportation agencies to evaluate how the difference of individuals' departure places such as residential, retail, and commercial area affects on mode choice behaviors.

국문요약

일반적으로 교통수단선택의 분석은 이용자의 수단별 선호도를 효용함수로 구축하여 분석하고 있다. 본 연구에서는 교통수단(버스, 철도)에 대한 이용자 선호도를 이용자의 출발지에 대한 공간적 영향을 함께 고려한 수단선택모형을 구축하여 분석하였다. 기존 효용함수가 이용자의 출발지에 대한 공간적 영향을 고려하지 못함에 따라 통행시간과 요금 등이 같으면 출발지에 상관없이 수단선호도가 같아지는 단점을 보완함에 따라 향후 지역별 수단분담 전환에 따른 배출가스 저감 등 다양한 분석이 가능할 것으로 판단된다.

1. 서론

교통수단에 대한 이용자 선호도 조사는 교통수단에 대한 이용자 수요를 추정하기 위한 중요한 기초자료로 활용된다. 교통수단에 대한 선호도는 개인교통수단인 승용차를 포함하여 현재 운행되고 있는 대중교통수단인 버스, 지하철(열차), 택시, 항공기 등 수단에 대한 조사가 있으며 이를 현시조사(Revealed Preference)로 정의한다. 현재 운행되지 않지만 장래 도입될 수단에 대한 선호도 조사는 (Stated Preference)로 정의되며 도입될 수단의 특성, 운행정보 등을 이용자에게 제시하여 선호도를 조사한다. 선호도 조사는 일반적으로 설문지를 작성하여 특정지역인 버스터미널, 철도역, 공항 등 교통유발시설지역이나 가구통행실태조사를 통해서 수행된다. 선호도조사는 전수조사가 아닌 표본조사이기 때문에 조사 응답자의 모든 출발지 또는 거주지에 대하여 대표성을 가지게 됨을 의미한다. 즉, 수단선호도를 나타내는 통행시간 및 비용이 유사하면 출발지의 지역적 특성과 관계없이 같은 수단분담률을 제시하게 된다.

†정회원, 한국철도기술연구원, 철도교통물류연구실
TEL : (031)460-5467 FAX : (031)460-5021
* 정회원, 한국철도기술연구원, 철도교통물류연구실
** 정회원, 고려대학교 의과대학 의학통계학교실
*** 정회원, 한국해양대학교, 데이터정보학과

지역적인 분포에 따른 영향은 교통네트워크에 의한 접근성을 대변할 수 있는 접근통행시간 등에 의해 단순히 고려되므로 이용자들의 평균적인 통행비용에 의해 수단선호도가 결정되는 구조를 가지고 있다.

본 연구에서는 이용자 표본조사자료를 기반으로 출발지 위치를 GIS기능을 활용하여 획득한 후 개별 응답자의 지역 간 교통수단인 버스와 철도에 대한 공간적 선호도 유무를 판단하고 수단분담률을 공간적 영향을 고려하여 추정하는 데 의의가 있다.

2. 공간모형의 고찰

2.1 공간자료

공간모형에서는 $\mathbf{s} \in D \subset \mathbf{R}^d$ 를 d 차원 유클리드 공간(Euclidean space) 상의 연구자의 관심지역 D 에서의 일반적인 자료의 위치 지점이라 하고, 공간 위치지점 \mathbf{s} 에서 관측 가능한 자료 $Y(\mathbf{s})$ 를 다음과 같은 다변량 확률장 (multivariate random field)로 정의한다.

$$\{Y(\mathbf{s}) : \mathbf{s} \in D\}$$

집합 D 의 속성에 따라서 $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ 는 지리통계자료, 지역자료, 점 패턴 자료와 같이 세 가지의 형태로 나누어진다(Cressie, 1993).

일반적으로 공간 자료는 특정 위치 또는 특정 지역에서의 관측 값이나 측정값으로 구성되며, 특히, 공간 자료는 지리통계자료(point-referenced data), 지역자료 (areal data), 공간 점 패턴자료 (spatial point patterns data)의 세 가지 유형으로 분류된다. 지리통계자료는 서울시 34개 측정기계에서 관측되는 미세 먼지 농도가 대표적인 예로써 측정기계가 없는 지역의 예측을 목적으로 한다. 이와는 대조적으로 지역자료는 서울시 25개 각 구에서 집계한 특정 질환자의 수가 대표적인 예로써 모형의 적합을 목적으로 한다. 지금까지의 연구들을 살펴보면 지리통계 자료에 대한 분석 및 연구는 많은 반면에 상대적으로 지역자료에 대한 연구는 다소 미진하다. 이에 본 논문에서는 실제 자료가 지역자료에 적합하므로 지역자료의 모형을 중심으로 다루고자 한다.

2.2 공간로지스틱회귀모형

본 절에서는 교통수단(열차와 버스)에 대한 선호도를 분석하기 위한 이항분포(Binomial distribution) 기반의 공간로지스틱회귀모형 (spatial logistic regression model)을 소개하며, 베이지안 (Bayesian) 방법인 마코프 연쇄 몬테칼로 (Markov chain Monte Carlo) 기법을 이용하여 공간로지스틱회귀모형 안에 포함되어 있는 모수를 추정하였다.

공간로지스틱회귀모형에 대한 간략한 설명은 다음과 같다. 서울시의 25개 구별로 열차와 버스의 교통수단 선택에 대한 승객들의 선호도에 대한 공간적 예측을 위하여 이항분포모형을 제시하였으며, 공간상 관성을 모형에 포함하는 모형으로 만들기 위해 제시한 공간로지스틱회귀모형은 아래와 같다.

$$Y(\mathbf{s}_i) \sim B(n_i, \pi(\mathbf{s}_i)) \quad (1)$$

$$\log\left(\frac{\pi(\mathbf{s}_i)}{1 - \pi(\mathbf{s}_i)}\right) = \mathbf{x}(\mathbf{s}_i)^T \boldsymbol{\beta} + \nu_\rho(\mathbf{s}_i)$$

여기서 B 는 성공의 확률(열차를 선택할 확률), $\pi(\cdot)$ 을 갖는 이항분포를 나타내고 $Y(\mathbf{s}_i)$ 는 i 번째 지역(구)에서 발생한 관심 사건의 회수(열차를 이용한 경우)이며, 가산자료(count data)의 형태로 나타난다. 또한 $\mathbf{x}(\mathbf{s}_i)$ 는 i 번째 지역(구)에 대한 설명변수의 열벡터(column vector)이고 $\nu_\rho(\mathbf{s}_i)$ 는 공간영향력을 나타내는 임의효과(random effect)를 의미하며 ρ 는 공간영향력을 나타내는 모수로서 일반적으로 조건부자기회귀모형 (conditional autoregressive model)으로 구성된다.

2.3 조건부자기회귀모형

본 연구에서 이용되는 지역별 교통수단 선호도 자료는 공간자료 중 지역자료에 해당되므로 지역자료를 위한 대표적인 모형인 조건부자기회귀모형에 대하여 간단히 살펴보면 다음과 같다.

공간상에서 관측되어진 $\{Y(\mathbf{s}_i); i = 1, \dots, n\}$ 가 정규확률 과정(Gaussian stochastic process)을 따르고 $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ 이 집합 D 의 격자형태를 가진다고 가정한다. 즉, 다음의 조건을 만족한다고 가정한다.

$$\bigcup_{i=1}^n \mathbf{s}_i = D, \quad \mathbf{s}_i \cap \mathbf{s}_j = \emptyset, \quad \forall i \neq j$$

그러면 $\{Y(\mathbf{s}_i); i = 1, \dots, n\}$ 을 아래와 같은 조건부공간자기회귀모형으로 표현할 수 있다.

$$Y(\mathbf{s}_i) | \{Y(\mathbf{s}_{-i})\} \sim N\left(\sum_{k=1}^n w_{ik} Y(\mathbf{s}_k), \tau_i^2\right)$$

여기서, N 은 정규분포를 나타내며, $\{Y(\mathbf{s}_{-i})\} = \{Y(\mathbf{s}_k); k \neq i\}$, $w_{ii} = 0$ 이고 $w_{ik}\tau_k^2 = w_{ki}\tau_i^2$ 이며, τ_i^2 는 i 번째 지역(구)에서의 조건부 분산 (conditional variance)을 나타낸다. 공간영향력을 모형에 포함하기 위해 필요한 이웃행렬 (neighborhood matrix), $\mathbf{W} = \{w_{ij}\}$ 은 0 또는 1의 값을 가지는 원소들로 구성된다. 즉, 서로 두 지역이 이웃하고 있으면 $w_{ij} = 1$, 그렇지 않으면 $w_{ij} = 0$ 을 가지는 이웃정보를 구성하고 조건부 분산은 $\tau_i^2 = \sigma_c^2 / w_{i+}$ 로 표현될 수 있다. 여기서, σ_c^2 은 공통분산을 나타내고 $w_{i+} \equiv \sum_{j=1}^n w_{ij}$ 는 i 번째 지역(구)에 이웃한 지역들의 수를 의미한다.

이웃행렬, $\mathbf{W} = \{w_{ij}\}$ 와 조건부 분산으로 이루어진 행렬, $\mathbf{T} = \text{diag}\{\tau_1^2, \tau_2^2, \dots, \tau_n^2\}$ 을 이용하여 인수분해정리에 의한 $\mathbf{y} = (Y(\mathbf{s}_1), Y(\mathbf{s}_2), \dots, Y(\mathbf{s}_n))^T$ 의 결합 확률 분포 (joint probability distribution)는 $\mathbf{y} \sim N(\mathbf{0}, (\mathbf{I}_n - \mathbf{W})^{-1}\mathbf{T})$ 로 표현될 수 있다 (Besag, 1974). 여기서 $\mathbf{I}_n - \mathbf{W}$ 는 역행렬이 존재하고, $(\mathbf{I}_n - \mathbf{W})^{-1}\mathbf{T}$ 는 대칭(symmetric)인 양정치 행렬(positive-definite matrix)이고 \mathbf{I}_n 은 n 차원 단위행렬을 나타낸다. 또한 공간 상관성을 나타내는 모수 ρ 를 고려한 조건부자기회귀모형은 아래와 같이 정의할 수 있다.

$$Y(\mathbf{s}_i) | \{Y(\mathbf{s}_{-i})\} \sim N\left(\sum_{k=1}^n w_{ik} Y(\mathbf{s}_k), \tau_i^2\right) \quad (2)$$

이 때 $Y(\mathbf{s}_i) \sim \text{CAR}(\rho)$ 라고 표현하며 식 (2)는 인수분해 정리에 의하여 식 (3)과 같이 나타낼 수 있다.

$$\mathbf{y} \sim N(\mathbf{0}, (\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{T}) \quad (3)$$

이러한 조건부자기회귀모형이 식 (1)에서 공간영향력을 나타내는 임의효과인 $\nu_\rho(\mathbf{s}_i)$ 에 대한 모형으로 적용할 수 있다.

2.4 모형 추정 방법

본 연구에서 제안된 공간로지스틱회귀모형의 모수 추정을 위하여 베이저안 기법을 이용하였으며 식 (3)에서 제시된 모형을 적합하기 위하여 모수들에 대한 사전분포 (prior distribution)에 대한 정의가 선행되어야 한다. 본 연구에서는 주요 모수들에 대한 사전분포들은 아래와 같은 무정보 사전분포 (non-informative prior distribution)를 갖는 접근방법을 이용하였다 (Hughes-Oliver et al., 2009).

$$\begin{aligned} \beta_p &\sim N(0, 10^3), p=1, \dots, 4, & \nu_\rho(\mathbf{s}_i) &\sim \text{CAR}(\rho), \\ \sigma_c^2 &\sim \text{gamma}(10^{-3}, 10^{-3}), \\ \rho &\sim \text{uni form}(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}) \end{aligned}$$

여기서, λ_{\min} 과 λ_{\max} 은 식 (5)에서의 $\mathbf{T}^{-1/2}\mathbf{W}\mathbf{T}^{-1/2}$ 의 최소, 최대 고유값(eigen value)을 의미하고, 공간상관모수인 ρ 는 이들 고유값의 역수를 범위로 가지게 된다.

본 연구에서 공간상관성을 반영하지 않은 일반적인 로지스틱회귀모형과 공간상관성을 반영한 공간로지스틱회귀모형을 제시하였으며, 각 모형별로 모수들을 추정하고 최적 모형을 선택하였다.

3. 실증분석

3.1 조사자료 및 모형

본 연구에 이용된 자료는 경부고속철도 개통이전인 2001년도에 고속철도개통시의 교통수단선택 변화를 분석하기 위해 조사된 선호의식 조사 자료를 이용하였다. 선호의식 조사는 승용차 이용자를 포함하여 주요 열차역, 버스터미널, 공항 등을 이용한 통행자를 대상으로 하였다. 본 연구에서는 서울시에서 출발하여 경상권, 전라권, 충청권을 통행하는 표본에 대하여 출발지를 기준으로 철도에 대한 선호도를 예측하였다.

본 모형에서 사용된 설명변수로는 평균 소득, 최종목적지까지의 총거리와 총비용을 고려하였으며 공간상관성을 고려하지 않은 일반적인 로지스틱모형과 공간로지스틱회귀모형을 비교 분석하였다.

모형 1. 일반 로지스틱회귀모형

$$\begin{aligned} Y(\mathbf{s}_i) &\sim B(n_i, \pi(\mathbf{s}_i)) \\ \log\left(\frac{\pi(\mathbf{s}_i)}{1-\pi(\mathbf{s}_i)}\right) &= \mathbf{x}(\mathbf{s}_i)^T\boldsymbol{\beta} \end{aligned}$$

모형 2. 공간 로지스틱회귀모형 (식 (3) 참조)

2.2절에서 소개한 공간로지스틱회귀모형은 일반 로지스틱회귀모형에서의, 관찰개체(본 연구에서는 지역(구)) 간의 독립성 조건을 가정하지 않는다. 즉, 관찰개체 간의 상호연관성을 가정한다. 특히, 특정 지역과 그 지역의 인접 지역들(이웃) 간에 수단선택은 서로 연관되어 있을 것으로 가정하고 지역들 간의 연관성 정도를 나타내는 구조를 일반 로지스틱모형에 포함한 형태가 바로 공간로지스틱회귀모형이다. 따라서 모형 2. 하에서 본 연구에서 고려한 설명변수들 뿐만 아니라 공간적 상호연관성의 효과도 특정 지역의 수단선택에 영향을 주는 지를 파악할 수 있다.

3.2 모수추정결과

본 연구에서 제시한 공간로지스틱회귀모형의 모수를 추정하기 위하여 Tanner and Wong(1987), Gelfand and Smith(1990)에 의해서 소개된 마코프 연쇄 몬테칼로 방법을 사용하여 범용 베이저안 통계 패키지인 WinBUGS (Spiegelhalter et al., 2000)를 사용하였으며, 추정된 모수 값의 수렴상태를 알아보기 위하여 세 가지의 서로 다른 초기값을 이용하였으며, 사후분포(posterior distribution)를 통해 모수 추정값(posterior mean; 사후평균)을 얻을 수 있었다. 추정된 모수의 올바른 수렴 여부를 확인하기 위해 Gelman and Rubin (1992)에 의해 제안된 방법을 이용하였다. 이 방법은 여러 개의 초기값을 가지고 수렴여부를 진단하는 것으로써 척도값이 1에 가까워지면 모든 초기값에 의해 생성된 모수들이 동일한 분포로 수렴한다는 것을 의미한다. 특정 지역의 이웃은 지역 구분 경계선을 공유하는 지역들로 정의한다.

본 연구에서는 제시하지 않았지만 공간로지스틱회귀모형에 대하여 세 개의 초기치를 적용한 결과 모수들의 사후분포 표본들에 대한 Gelman-Rubin 수렴진단 척도값이 1에 가깝게 나타나고 있으므로 생

성된 모수들이 잘 수렴하여 올바른 추정 값을 확인하였다. <표 1>은 공간상관성을 반영하지 않은 모형(모형 1)하에서의 추정값들을 나타낸다. 여기서, β_1 는 상수항, β_2 은 평균소득에 관한 계수, β_3 는 총거리
리에 관한 계수, β_4 은 총비용에 관한 계수를 의미한다.

Table 1. Parameter estimates and confidence intervals from Model 1

	일반 로지스틱회귀모형 (모형 1.)	
	추정값	95%신뢰구간
상수항	0.340	-194.4 ~ 194.0
평균소득	0.013	-0.222 ~ 0.248
총거리	-1.291	-1.709 ~ -0.888
총비용	1.142	0.716 ~ 1.581

표 1.을 통해 알 수 있듯이 평균소득은 열차선호도에 영향을 미치지 않는 것으로 나타난 반면에 총거리와 총비용은 열차 선호도에 영향을 미치는 것으로 나타났다. 구체적으로는 총비용은 열차선호도에 정(+)
의 영향을 미치는 반면에 총거리는 부(-)의 영향을 미치는 것으로 나타났다. 즉, 총거리의 경우 거리가 증가할수록 열차를 선호할 확률이 낮아지는 반면에 총비용이 증가할수록 열차를 선호할 확률이 높아
짐을 알 수 있다.

Table 2. Parameter estimates and confidence intervals from Model 2

	공간 로지스틱회귀모형 (모형 2.)	
	추정값	95%신뢰구간
상수항	0.389	0.147 ~ 0.658
평균소득	0.047	-0.239 ~ 0.330
총거리	-1.487	-1.995 ~ -1.011
총비용	1.289	0.789 ~ 1.828
공간상관성	-0.003	-0.046 ~ 0.031
공통분산	0.479	0.047 ~ 1.119

표 2.는 공간상관성을 반영한 모형에 대한 추정값을 나타낸다. ρ 는 음의 추정값으로 나타났으나 95%신뢰구간 안에 0이 포함됨에 따라 통계적으로는 유의하지 않은 것으로 나타났다.

공간상관성을 반영하지 않은 모형의 추정값(표 1.)과 비교해 보면 역시 평균소득은 통계적으로 유의하지 않은 반면, 총거리와 총비용은 유의한 것으로 나타났다.

3.3 DIC와 MSPE 관점에서의 모형 비교

본 연구에서는 공간상관성을 반영하지 모형과 반영한 모형을 비교하기 위한 기준들로 DIC (Deviance Information Criterion)와 MSPE (Mean Squared Prediction Error)를 고려하였다.

주어진 모형의 적합도를 평가하는 데 있어서 항상 문제가 되는 점은 모형이 복잡해질수록 항상 자료를 더 잘 적합할 수 있다는 것이다. 한편, 모형의 설명 능력은 우도함수의 값으로 측정될 수 있으나 복잡한 모형보다는 간단한 모형이 선호되어야 한다는 모수 절약의 원리에 의하여 추정되어질 모수의 수가 많은 것에 대해 벌칙(penalty)을 부가함으로써 모형의 절약성을 고려할 수 있다. 베이지안 기법에서 모형의 설명력과 절약성은 서로 상충되는 기준으로서 여러 가지 모형 선택의 판단기준을 함께 고려한 선택 기준이 필요하게 되는데 DIC 기준을 이용하였다.

베이지안 적합도 기준인 DIC는 일반적인 적합도 기준인 AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion)와 같이 모형의 복잡성을 고려한 적합도 기준이다.

모형 적합도는 $D(\theta) = 2\log[f(\mathbf{y}|\theta)] + 2\log[h(\mathbf{y})]$ 의 사후분포 평균인 $\bar{D} = E_{\theta|\mathbf{y}}[D(\theta)]$ 이며 $h(\mathbf{y})$ 는 모형에

의하여 영향을 받지 않는다. $\theta = (\beta_1, \beta_2, \beta_3, \beta_4, \rho, \sigma_c^2)$ 이고 모형의 복잡성은 모형에 영향을 주는 모수의 갯수인 $p_D = \bar{D} - D(\hat{\theta})$ 으로 측정되어질 수 있다. 여기서 $D(\hat{\theta})$ 는 모수의 사후평균, $\hat{\theta}$ 에서 평가된 데비언스의 값이다. 결과적으로 $DIC = \bar{D} + p_D$ 로 정의되며 작은 값일수록 모형 적합이 잘 되었다고 할 수 있다.

교통수단 선호도 모형을 위하여 제시된 두 모형에 대하여 가장 적합한 모형을 선택하기 위하여 앞서 설명한 DIC와 아래와 같이 정의되는 MSPE를 통하여 최적 모형을 선택하였다. 물론, MSPE는 DIC와는 달리 모형에 포함된 모수의 수에 대한 벌칙은 고려되지 않지만 DIC와 마찬가지로 모형과 실제자료와의 괴리 정도를 측정하는 적합기준으로 많이 사용된다. DIC와 MSPE 척도는 값이 작을수록 모형이 주어진 자료에 잘 적합되었다고 할 수 있다.

$$MSPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y(\mathbf{s}_i)}{n_i} - \hat{\pi}(\mathbf{s}_i) \right)^2$$

여기서, $\hat{\pi}(\mathbf{s}_i)$ 는 각 모형 하에서 구한 i 번째 지역(구)에서 열차를 선택할 사후확률(posterior probability 혹은 posterior proportion)값이다.

Table 3. Diagnostic measures from Model 1 and Model 2

	일반 로지스틱		공간 로지스틱	
	추정값	95%신뢰구간	추정값	95%신뢰구간
DIC	116.9		106.5	
MSPE	0.017	0.015 ~ 0.022	0.010	0.006 ~ 0.016

표 3.은 모형 비교를 위한 DIC와 MSPE 척도의 값을 제시하였다. 먼저, MSPE 관점에서는 공간 영향을 고려하지 않은 모형 1(일반 로지스틱 회귀모형)보다 모형 2(공간 로지스틱 회귀모형)가 훨씬 더 좋은 적합 정도를 보이고 있다. 하지만 이는 모형 2에 포함된 모수가 모형 1에 비해 많음으로 인해 우연히 발생된 결과일 수 있다. 따라서 모수의 개수에 대한 벌칙을 고려하는 적합기준인 DIC를 각 모형에 적용한 결과 모형 1이 공간적 상관구조를 고려한 모형 2에 비해 주어진 자료를 설명하는 데 다소 부적합하다는 판단을 내리게 된다.

3.4 열차 선호도에 대한 사후확률 지도

그림 1.은 서울시 25개구의 지역 간 통행자의 열차 선호도에 대한 실제 확률(=열차를 이용한 통행자 수/전체 통행자 수)을 나타낸 그림이다. 관측된 실제 비율을 모형 1과 모형 2를 적용한 결과 그림 2.와 그림 3.과 같은 사후확률에 대한 지도를 구축하였다. 그림 2.는 공간상관성을 반영하지 않은, 획일화된 평균적인 수단 선호도에 바탕을 둔 일반 로지스틱모형(모형 1)에 의해 얻어진 열차선호도에 대한 사후확률지도이고 그림 3.은 공간상관성을 반영한, 즉 모형 1과는 다소 차별화된 공간 로지스틱회귀모형(모형 2)으로부터 얻어진 열차선호도에 대한 사후확률의 결과를 나타내고 있다.

서울시에서 출발한 지역 간 통행자중 열차를 이용한 사람들에 대한 선호도 조사결과를 서울시 구별단위로 확률을 제시한 결과 그림 1.에서 보는 바와 같이 강북지역이 강남지역에 비해 열차 선호도가 높은 것으로 나타났으며 강남지역의 경우 동작구, 관악구, 서초구, 강남구의 열차 선호도는 상대적으로 낮은 것으로 분석되었다.

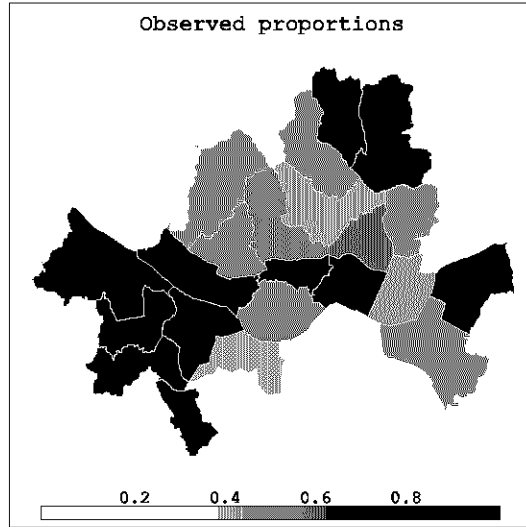


Fig 1. Observed probability map for train preference

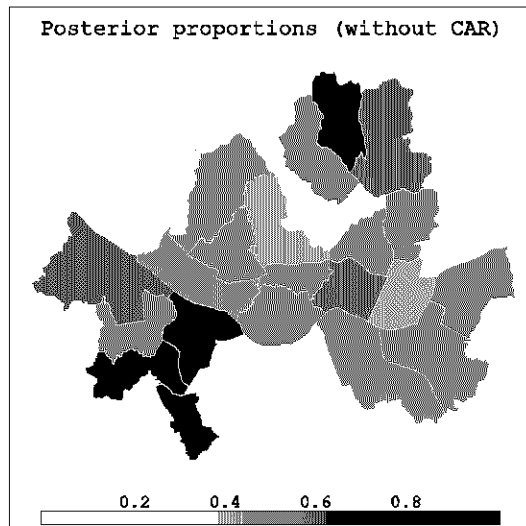


Fig 2. Posterior probability map for train preference from Model 1

모형의 비교를 위해서 열차수단 선호도의 공간상관성을 반영하지 않은 모형(그림 2.)과 반영한 모형(그림 3.)을 그림 1.과 비교해 보면 열차 선호도가 공간상관성을 반영하지 않은 로지스틱모형보다 공간상관성을 반영한 로지스틱모형에서 얻은 결과와 유사함을 알 수 있다.

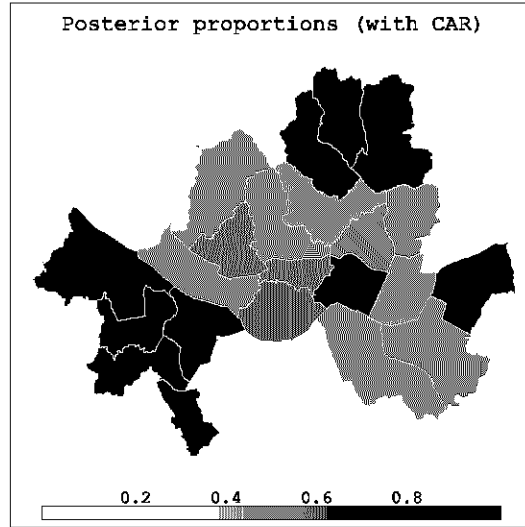


Fig 3. Posterior probability map for train preference from Model 2

이는 그림 4.를 통해서 보다 분명해 진다. 실제 확률(Obs)에서 각 모형(M.1, M.2)에서 얻은 사후확률을 뺀 차이값(Obs-M.1, Obs-M.2)을 이용하여 구성된 상자그림을 보면 일반 로지스틱회귀모형(M.1) 하에서의 사후확률은 전체적으로 과소추정된 경향을 보이고 있으나 공간 로지스틱회귀모형(M.2)은 차이값의 분포가 0을 중심으로 대칭적인 형태를 취하고 있으므로 안정적인 적합이 이루어졌음을 알 수 있다. 또한 차이의 절대값(|Obs-M.1|, |Obs-M.2|)을 비교한 결과 모형 1(M.1)의 변동이 모형 2(M.2)에 비해 훨씬 심한 것을 쉽게 파악할 수 있다. 따라서 공간 상관성을 반영한 모형을 통해 통행자의 수단 선호도를 예측하는 것이 보다 정확하고 신뢰성있는 결과를 도출하는 것으로 판단된다.

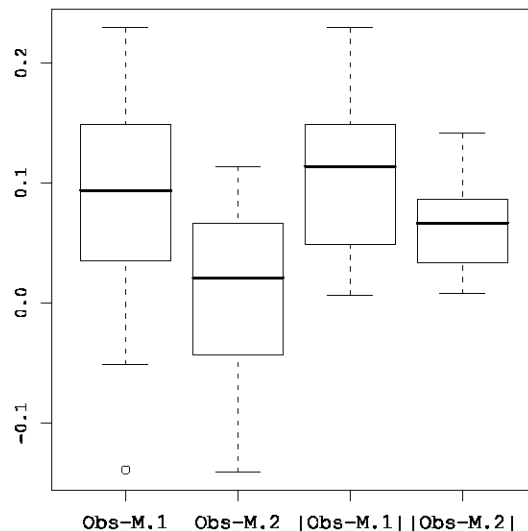


Fig 4. Box plot for comparison between observed probability and posterior probability from Model 1 and 2

특정 수단의 선택에 있어서 설명변수의 효과와 더불어 인접지역 간의 유사성의 효과를 모형에 포함시킴으로써 보다 신뢰성 있는 예측 및 적합이 이루어졌음을 각 모형에서 얻은 사후확률과 실제 확률을 비교해 봄으로써 알 수 있다.

4. 결론

본 연구에서는 지역 간 교통수단인 철도에 대하여 출발지(구단위)를 중심으로 공간상관성을 반영한 새로운 교통수단선택 모형을 구축하였다. 출발지의 공간상관성을 반영한 모형은 그렇지 않은 모형에 비해 수단선택 확률을 보다 정확하게 추정하는 것으로 분석되었으며 특히 지역적 특성을 반영할 수 있는 장점을 가지는 것으로 분석되었다.

공간상관성을 반영한 수단선택 모형은 향후 철도시설과 같이 특정 지역에 신규서비스가 이루어질 경우 이에 따른 공간적 영향을 반영한 수단분담의 추정이 가능하므로 향후 활용도가 높을 것으로 분석된다. 또한 대중교통 우선정책에 따라 여러 가지 정책적인 변수에 의해 타 수단으로부터 열차 수단으로 전환이 될 경우(modal shift) 수단 전환에 따른 환경개선효과(배기가스 절감효과) 등을 지역적으로 분석하는데 유용하게 이용될 것으로 판단된다.

본 연구에서는 한정된 크기의 표본으로 인해 서울시 25개 구 단위로 모형을 구축하여 분석하였으나 보다 많은 자료를 이용할 경우 서울시의 540여개 동단위로 수단선택도의 분석이 가능할 것으로 판단된다. 따라서 향후 연구과제는 공간상관성을 반영한 모형의 다양한 적용을 통한 활용방안의 검토와 다양한 수단의 영향을 분석할 수 있는 방법론의 구축이 필요할 것으로 판단된다.

참고문헌

1. 로짓·프라빗모형 응용, 박영사 (2005).
2. 경부고속철도 연계교통체계 구축 기본계획 수립 연구, 한국철도기술연구원 (2003).
3. 윤대식, 육태숙, 윤성순(2005) 경부 고속철도 2단계 개통후 대구시민의 지역간 통행수단 선택행태와 전환수요 예측에 관한 연구, 대한국토·도시계획학회지, 제40권 제5호, pp.113-119.
4. Haider, M. and E.J. Miller, "Effects of Transportation Infrastructure and Location on Residential Real Estate Values: Application of Spatial Autoregressive Techniques," Transportation Research Record 1722, TRB, National Research Council, Washington, D.C., 2000, pp. 1-8.
5. Hughes-Oliver, J. M., Heo, T-Y, Ghosh, S. (2009), An Autoregressive point source model for spatial processes, accepted in Environmetrics.
6. Fotheringham, A.S., C. Brunson, and M. Charlton, Geographically Weighted Regression: the analysis of spatially varying relationships, John Wiley & Sons Ltd., West Sussex, England, 2002.
7. Gelfand, A. E. and Smith, A. F. M. (1990) Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association, Vol. 85, pp. 398-409.
8. Gelman, A. and Rubin, D. B. (1992), Inference from iterative simulation using multiple sequence (with Discussion), Statistical Science, Vol. 7, 457-511.
9. Tanner, M. A. and Wong, W. H. (1987), The calculation of posterior distributions by data augmentation (with discussion), Journal of the American Statistical Association, Vol. 82, pp. 528-550.
10. Varameth V, Kazuaki M, Yoshiyuki T(2005) Mixed Logit Model Framework with Structuralized Spatial Effects: A Test of Applicability with Area Unit Systems in Location Analysis, Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 3789-3802.
11. Ben-Akiva, M., Steven R. Lerman (1987) Discrete Choice Analysis: Theory and application to travel demand, MIT Press.
12. Ben-Akiva, M., Watanatada (1981) Application of a Continuous Choice Logit Model: In structural analysis of discrete data with econometric applications, MIT Press.
13. Spiegelhalter, D. J., Thomas, A., and Best, N. G. (2000), WinBUGS Version 1.3 User Manual. Cambridge: Medical Research Council Biostatistics Unit.