

수단선택의 이질성을 고려한 잠재계층모형(Latent Class Model) 구축: 서울시 통근자를 사례로

Developing a Latent Class Model Considering Heterogeneity in Mode Choice Behavior : A Case of Commuters in Seoul

김 성 후* · 추 상 호**

* 주저자 : 조지아공과대학 토목공학과 박사과정

** 교신저자 : 홍익대학교 건설도시공학부 교수

Sung Hoo Kim* · Sangho Choo**

* School of Civil and Environmental Engineering, Georgia Institute of Technology, USA

** Dept. of Urban Design & Planning, Hongik University

† Corresponding author : Sangho Choo, shchoo@hongik.ac.kr

Vol.18 No.2(2019)

April, 2019

pp.44~57

pISSN 1738-0774

eISSN 2384-1729

<https://doi.org/10.12815/kits.2019.18.2.44>

2019.18.2.44

Received 19 March 2019

Revised 5 April 2019

Accepted 8 April 2019

© 2019. The Korea Institute of Intelligent Transport Systems. All rights reserved.

요 약

교통계획에서 사람들의 수단선택 의사 결정과정을 이해하고 이를 토대로 수요예측의 정확성을 높이는 것은 매우 중요하다. 이를 위해 여러 계층마다 수단선택 영향요인들이 다르게 나타나는 이질성(taste heterogeneity)을 반영한 모형들이 개발되어 왔다. 본 연구에서는 목적지의 토지이용 특성에 따라 수단선택에 미치는 요인의 영향력이 다를 것이라는 가설을 세우고 2010년 서울시 가구통행실태조사 자료를 활용하여 서울시내 출근통행에 대한 잠재계층모형을 분석하였다. 먼저 목적지의 토지이용 특성을 활용하여 수단선택에 대한 잠재계층을 구분하고, 각각의 잠재계층에 대한 수단선택 모형을 개발하였다. 잠재계층모형 추정결과, 서울시내 출근통행 수단선택의 경우 두 개의 잠재계층으로 분석되었다. 첫 번째 계층은 승용차와 대중교통의 수단분담이 비슷하고, 목적지가 상대적으로 도심이 아닌 주거시설이 많은 지역들이고, 두 번째 계층은 주로 대중교통의 분담률이 높고 목적지가 상대적으로 도심에 속하는 업무/상업시설이 많은 지역이다. 또한 이들 계층간의 수단선택 모형의 경우, 인구 및 사회경제지표들의 수단선택에 미치는 영향이 계층간에 차이가 있는 것으로 분석되었다.

핵심어 : 이질성, 수단선택, 잠재계층모형, 가구통행실태조사, 통근통행

ABSTRACT

It is crucial to understand how people make decisions on mode choice and to accurately predict their behaviors in transportation planning. One of avenues for advancing modeling is, in particular, taking into account for taste heterogeneity in modeling that can incorporate different decision-making processes across group. In this study, we hypothesize that how people make decisions on mode choice would differ by destination in that land use characteristics are heterogeneous by zone even if zones are all in the same area. To this end, we apply Latent Class Modeling (LCM) to commute trips in Seoul by using 2010 household travel diary survey, investigate types of latent classes with the aid of characteristics of destination, and analyze how those classes differently response to factors. The LCM identifies two classes: in the first one, modal split of auto and public transit (bus and metro) is almost half-and-half and the trip destinations are characterized by relatively more residence facilities and less business/commercial facilities; in the second one, public transit has a notably high share and

trip destinations are characterized by relatively more business/commercial facilities. In addition, it turns out that demographic and socio-economic variables affect mode choice differently by class.

Key words : Taste heterogeneity, Mode choice, Latent class model, Household travel survey, Commute trips

I. 서론

서울시 대중교통은 2007년 대중교통 시스템 개편을 통해 수도권에서 자가용과 함께 주요 통행수단으로 자리매김하고 있다. 이들 통행수단과 각각의 분담 역할은 각종 교통정책과 도시계획에 있어서 핵심 요소이다. 따라서 사람들의 수단선택의 의사결정과정을 이해하고 정확하게 이용을 예측하는 것은 매우 중요하다. 이를 위해 교통수요모형 분야에서는 기존의 수단선택모형에 현실성을 좀더 반영하여 개선하려는 노력들이 다양한 방법으로 행해져 왔다.

먼저 통행행태를 설명하기 위한 설명변수로서 대부분의 연구가 개인의 사회경제지표나 선택대안의 성격이 이용되어 왔으나, 추가적으로 개인의 성향(Choo and Mokhtarian, 2004; Johansson et al., 2006; Lanzini and Khan, 2017) 혹은 주변의 환경적 특성(Handy et al., 2005; Sun et al., 2016)에 대한 고려가 중요하다는 것이 여러 연구를 통해 밝혀져 왔다. 모형적인 측면에서는 기존 모형의 제약을 줄이는 방향으로 많은 모형이 개발되어 왔다. 예를 들면, 선택대안간의 비관측 상관관계를 고려하기 위해 다양한 형태의 네스티드 로짓(nested logit) 모형이 이용되었고(Carrasco and Ortuzar, 2002), 개인의 잠재성향을 잠재변수로서 활용하기 위해 하이브리드 선택(hybrid choice) 모형이 개발되었다(Ben-Akiva et al., 2002; Vij and Walker, 2016). 또 다른 방법으로 선택모형에 taste heterogeneity(또는 taste variation, 이하 이질성)을 도입한 혼합로짓(mixed logit) 모형과 잠재계층모형(latent class model) 등이 활용되었다(Hensher and Greene, 2003; Greene and Hensher, 2003). 여기서 이질성은 의사결정자 혹은 의사결정자들의 집단마다 선택에 영향을 미치는 요소들에 대해 다르게 중요도를 매기고 반응한다는 것이다. 실제로 선택에 대한 이질성이 유의미하게 존재한다면, 단순형 모형처럼 이질성을 반영하지 못할 경우 통행행태에 대한 예측력이 떨어지고 모형 파라미터(parameter)의 편이(bias)를 야기시킨다. 이를 극복하기 위해 마케팅 분야에서는 모집단을 몇몇 그룹으로 나누고(market segmentation), 이들 그룹 간의 차이에 대한 이해를 바탕으로 타겟(target) 그룹에 따라 서로 다른 마케팅 전략을 사용하고 있다. 이와 마찬가지로 교통에서도 수단선택에 있어 선호가 다른 계층들을 알 수 있다면, 정책 시행에 있어 그룹별 맞춤형 전략을 세우는데 더 도움이 될 수 있다.

본 연구에서는 국내에서 상대적으로 연구가 부족했던 변수인 지역의 특성 혹은 토지이용 변수를 활용하며, 수단선택모형을 구축하는데 있어 이질성을 고려하고자 한다. 특히, 서울시 내부의 경우 토지이용과 건축 물용도가 지역에 따라 매우 상이하기 때문에 통행의 목적지가 어디냐에 따라 통행행태가 변할 수 있다. 따라서 통행 목적지의 특성과 수단선택의 잠재계층을 연결시킨 잠재계층모형을 구축함으로써 보다 현실적인 수단선택행태를 반영하고자 한다.

II. 선행연구검토

선택모형의 기술적 관점에서 이질성(taste heterogeneity)이란 행태 및 의사결정 과정 모형에서 파라미터의 값이 모집단 내에서 다를 수 있다는 것이다. 즉, 기존의 모형들은 파라미터를 추정할 때 모집단 전체의 평균

을 추정하는 반면, 이질성을 고려한 모형은 파라미터의 분포를 추정하거나 여러 계층 각각의 평균을 추정하는 것이다. Kim and Mokhtarian(2018)은 선택모형에서 이런 이질성을 고려하는 접근법에 대한 유형분류(typology)를 하였는데, 계층을 분류하는 방법에 따라 비확률적 분류와 확률적 분류로 나눌 수 있고, 파라미터 분포에 대해 가정하는 방법에 따라 연속성과 불연속성으로 구분하였다. 대표적인 모형이 혼합로짓모형(파라미터 분포의 연속형을 가정한 확률적 분류)과 잠재계층모형(파라미터 분포의 불연속성을 가정한 확률적 분류)이다. 여러 가지 방법론에 대한 비교연구도 진행되어 왔는데, 대표적으로 Greene and Hensher(2003)는 잠재계층모형과 혼합로짓모형을 이론적 측면에서 비교하고 실증자료에 적용하였다. 이 연구에서는 잠재계층이 혼합로짓에 비해 모형의 적합도가 다소 높은 것으로 나타났으나, 어느 모형이 더 우수하다기 보다는 각각의 장점이 있음을 강조하였다. 혼합로짓 모형의 경우 파라미터의 분포를 추정하여 모형구축의 융통성이 높으나, 여러 가정이 필요한 단점(파라미터의 분포를 갖는 변수가 무엇인지, 어떤 분포를 갖는지, 파라미터간의 상관관계 등)이 지적되어 왔다(Vij et al., 2013). 잠재계층모형은 잠재계층이 얼마나 어떻게 존재하는지 사전에 알 수 없다는 단점이 있으나, 그룹을 나누고 각 특성을 분석할 수 있다는 측면에서 마케팅 분야에 많이 활용되어 왔으며, 교통분야에서도 보다 현실적인 선택모형을 위하여 종종 이용되어 왔다.

수단선택과 관련해서 Bhat(1997)는 잠재계층모형을 도시간 이동 수단선택(자가용, 기차, 비행기)에 적용하였으며, 모형추정 결과 3개의 잠재계층으로 분류되었으며 각 계층들의 주요 변수들이 서로 다른 민감도를 가지고 있음을 확인하였다. Vij et al.(2013)과 Vij and Walker(2014) 또한 독일 도시와 미국 샌프란시스코 주변의 수단선택에 대한 모형을 적용하여 각각 3개와 6개의 잠재계층을 도출하였다. 특히 Vij and Walker(2014)는 기존의 모형과 비교하여 잠재계층은 통행행태를 이해하는데 더 많은 정보를 주며, 기존 모형의 경우 편이(bias)된 결과를 가져올 수 있다는 것을 입증하였다. 수단선택 외에도 잠재계층모형은 주거지역 선택(Walker and Li, 2007), 차량소유(Anwar et al., 2014; Kim and Mokhtarian, 2018) 등에도 적용되어 왔다. 특히 Kim and Mokhtarian(2018)은 이질성에 대한 고려가 내재적 편이(endogeneity bias)와 어떤 연관성이 있는지, 이를 해결하는데 있어 잠재계층모형이 어떻게 활용하는지에 대하여 보여주었다. 또한 잠재계층모형이 기존의 다항로짓(multinomial logit) 모형이나, 비확률적 모형에 비해 우수함을 보였다.

국내에서는 Bae et al.(2010; 2011)이 잠재선호를 바탕으로 수단선택에 대한 잠재계층모형을 개발하였지만, 다른 교통분야에 잠재계층모형을 활용한 연구는 매우 미미하다. 먼저 Bae et al.(2010)은 한강수상대중교통을 고려한 수단선택모형 추정에 잠재계층모형을 적용하였다. 잠재계층모형에는 한강수상대중교통 잠재이용자 설문조사결과중 수상교통과 대중교통이용 선호도를 반영하였으며, 수단선택모형은 개인속성, 통행시간, 통행비용 등을 고려하였다. 잠재계층 수단선택모형 분석결과, 잠재계층간의 수단선택에 영향을 주는 요인이 차이가 있는 것으로 나타나 대중교통의 심리적인 선호도에 따라 수단선택행위의 차이가 있음을 확인하였다. 또한, Bae et al.(2011)은 2006년 가구통행실태조사 자료를 활용하여 쇼핑목적 통행의 수단선택에 대한 잠재계층모형을 개발하였다. 먼저 통행 목적지의 아파트수, 종사자수, 인구밀도, 버스정류장수, 지하철역 수 등 토지이용변수를 잠재계층에 반영하였으며, 수단선택에서는 개인특성변수, 통행시간 및 통행비용 등을 고려하여 추정하였다. 모형추정결과 각 계층간에 수단선택에 영향을 주는 요인들이 다르게 나타나 계층간의 수단선택행동의 이질성이 존재하는 것으로 분석되었다.

한편, 통행행태 및 수요예측에 있어 지역적 특성과 토지이용의 성격은 매우 중요한 요소로 사용되어 왔다. Sun et al.(2016)은 상당수의 통근수단 관련연구가 주거지역(출발지)의 토지이용 특성에 초점을 맞췄으며, 직장 위치(목적지)의 토지이용의 영향에 대한 연구는 상대적으로 적은 것으로 분석하였다. 이러한 관점에서 본 연구는 목적지인 직장의 위치에 보다 중점을 두었다. 또한, 토지이용은 다양한 형태로 모형에 반영되었는데, 자주 이용되어온 변수는 'D' 변수들인 density, diversity, design, destination accessibility, distance to transit이다

(Cervero and Kockelman 1997; Ewing and Cervero, 2010). 그 외에도 목적지의 타입을 고려하기도 했는데, Popovich and Handy(2015)는 쇼핑목적지의 타입과 수단선택의 연관성에 대해 연구했다. 이 연구에서는 쇼핑 목적지를 downtown, strip center, big box로 구분하고, 각각에 대해 도보/자전거 선택여부에 대한 이항로짓(binary logit) 모형을 개발하였다. 한가지 주목할 점은 지역적 특성과 토지이용을 어떻게 모형에 반영하였는지이다. 대부분의 연구에서는 지역적 특성과 토지이용을 모형에 직접적인 설명변수로 포함한다. 즉, 그러한 특성들이 선택대안의 효용에 직접적인 영향을 미치는 것으로 가정하는 것이다. 가령 Reilly and Landis(2003)과 Zhang(2004)는 다양한 지역적 특성과 토지이용을 다항로짓모형(multinomial logit model)의 설명변수로 활용하였고, 또한 국가교통DB에서 제공하는 장래수요예측의 수단선택모형도 단순히 지역변수를 더미(dummy) 변수로 활용하였다(The Korea Transport Institute, 2012). 다른 방법은 Popovich and Handy(2015)처럼 지역별로 표본을 미리 구분하고 각각의 모형을 구축하는 것이다(앞서 언급한 비확률적 분류법과 같다).

결론적으로 지역적 특성/토지이용과 통행행태와의 연관성에 관한 연구에 있어서 (1) 목적지의 성격에 대한 연구, (2) 이질성에 대한 고려, (3) 특히 확률적 방법론이 적용된 연구는 부족했다. 따라서 본 연구는 선행 연구에서 부족했던 부분들을 보완하기 위해 잠재계층모형과 목적지의 특성변수들을 활용하여 그들의 활용도를 실증적으로 보여주는 것을 목표로 하였다.

Ⅲ. 분석자료 및 분석방법

1. 분석자료

본 연구에서는 서울시 출근통행에서의 수단선택의 이질성을 분석하기 위해 가구통행실태조사 자료를 활용하였다. 가구통행실태조사는 설문지 형태로 읍면동의 행정조직 등을 통해 배포되었으며, 응답자가 평일 하루 (목요일) 동안의 통행을 기록하도록 디자인되었다. 2016년 가구통행실태조사 자료의 구득이 가능하지 않아 본 연구에서는 이용가능한 최신자료인 2010년 가구통행실태조사 자료를 활용하였다. 데이터에는 가구정보, 가구원정보, 통행정보가 포함되어 있으며, 조사와 데이터에 대한 자세한 설명은 보고서를 참조하면 된다(The Korea Transport Institute, 2011). 특히, 본 연구에서는 도착지의 토지이용 특성을 반영하기 위해 한국토지정보시스템(KLIS, <http://klis.seoul.go.kr>)에서 제공하는 과세대상자료로부터 서울시내 존 (동 단위)에 대한 용도별 시설물 면적 데이터를 취득하여 활용하였다. 본 연구는 서울 시내에서의 출근통행에 초점을 맞췄으며, 유의미한 결과를 도출하기 위해 세 가지 주요 통행수단 (승용차, 버스, 지하철/전철)에 한정하였다. 승용차는 직접운전과 다른사람 운전을 포함하며, 버스는 시내버스와 광역버스를 포함한다. 여러 통행수단이 환승 등을 통해 이용될 수 있지만, 본 연구에서는 주 통행수단에 한정하여 연구하였다. 추가적으로 18세 이하와 데이터 누락이 있는 케이스를 제거하고 최종적으로 72,595개의 통행에 대하여 분석을 시행하였다.

최종분석 데이터에 대한 기초통계는 <Table 1> 및 <Table 2>와 같다. 수단선택의 경우 대중교통이 과반 이상을 차지하고 있고(지하철41%, 버스21%), 남성이 66%이며, 대다수가 운전면허(84%)나 가구내 차량을 보유(84%)하고 있다. 가구월소득의 경우 피라미드형으로 분포하고 있으며 절반이 2.5백만원 사이에 분포해 있다. 직업의 경우 사무직 직종이 36%를 차지하고 있으며, 다른 직종 또한 고르게 분포하고 있다. 통근자의 평균 연령은 42세이고, 집에서 지하철역, 버스정류장까지의 평균 도보시간은 각각 11분, 5분으로 나타났다. 본 연구에서 별도로 취득한 시설물 이용변수들의 경우, 목적지 존의 시설물 이용을 다섯가지로 분류하였으며(주거, 상업, 업무, 교육, 산업), 모두 표준편차가 상대적으로 큰 것으로 보아 존별로 시설물 이용이 상이함을 알

수 있다. <Fig. 1>은 목적지 존의 시설물 면적을 시설물 유형별로 나타낸 것이다. 특히 상업 및 업무시설이 특정지역에 모여 있음(도심)을 알 수 있다.

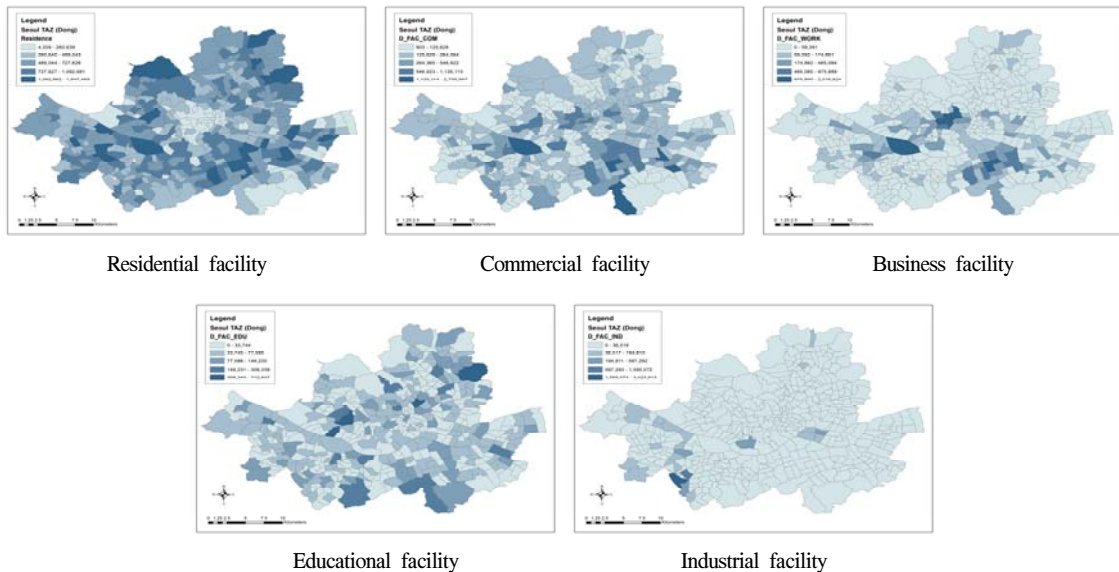
<Table 1> Descriptive statistics of key categorical variables

Variable	Category	Count	Percent
Commute mode	Auto	27,470	37.8%
	Bus	15,410	21.2%
	Metro	29,715	40.9%
Sex	Male	47,766	65.8%
	Female	24,829	34.2%
Driver licensure	Yes	60,833	83.8%
	No	11,762	16.2%
Car ownership	Yes	60,898	83.9%
	No	11,697	16.1%
Household monthly income	Less than 1 million won	3,518	4.8%
	1-2 million won	11,738	16.2%
	2-3 million won	19,244	26.5%
	3-5 million won	23,963	33.0%
	5-10 million won	12,386	17.1%
	10 million+ won	1,746	2.4%
Occupation	Professional	10,676	14.7%
	Service worker	10,006	13.8%
	Sales worker	10,057	13.9%
	White collar	26,522	36.5%
	Blue collar	6,829	9.4%
	Other	8,505	11.7%
Work schedule	5 days weekly	38,238	52.8%
	6 days weekly	18,709	25.8%
	5 days biweekly	4,140	5.7%
	Other	11,401	15.7%

변수이용에 있어 더 용이한 모형을 위해 몇몇 변수를 변환하였다. 연령의 경우 라이프사이클을 좀 더 현실적으로 반영하기 위해 특정 연령대로 구분하고(19-29, 30-49, 50-64, 65+), 소득은 3개의 수준으로 구분하고(2백만미만, 2-5백만, 5백만원이상), 직업의 경우 가장 루틴한 통근통행을 보일 것으로 보이는 사무직종과 그 외를 구분하여 더미변수로 활용하였다. 시설물 이용면적의 경우 수치가 상대적으로 크고 분포가 편향(skewed)되어 있어 스케일을 줄이고 정규분포화하기 위해 자연로그를 적용하였다. 모형구축에 있어 다양한 조합의 변수를 실험하였으며, 최종적으로 유의미한 결과를 도출하는 변수를 선택하여 최종모형을 구축하였다. 한편, 수단선택의 모형에 일반적으로 포함되는 통행시간과 통행비용의 경우 본 연구의 모형에는 포함되지 않았다. 이는 본 분석 자료에 선택되지 않은 대안들에 대한 통행시간과 통행비용에 대한 정확한 정보가 없기 때문이다. 또한, 본 연구의 주요 초점이 토지이용의 이질성과 수단선택의 관계이기 때문에 모형구축에 있어 이들 변수가 누락되어도 큰 무리가 없다고 판단하였다.

<Table 2> Descriptive statistics of key continuous variables

Variable	Mean	Standard deviation
Age	42.19	11.60
Number of household members	3.39	1.08
Number of children	0.15	0.44
Access time to metro station(min)	11.05	7.64
Access time to bus stop(min)	5.40	3.55
Residential facility of destination(square meters)	666,798.00	403,745.80
Commercial facility of destination(square meters)	417,260.00	393,219.80
Business facility of destination(square meters)	346,216.50	549,035.20
Educational facility of destination(square meters)	63,284.05	83,110.00
Industrial facility of destination(square meters)	74,878.85	428,459.10



<Fig. 1> Zone characteristics(facility area by type)

2. 분석방법

기준에 가장 많이 활용되는 수단선택 모형은 효용 극대화 이론(utility maximization theory)을 기반으로 하는 다항로짓(MNL; multinomial logit) 모형이다. 교통과 관련된 상당수의 연구가 다항로짓 모형을 기반으로 수행되어 왔다. 다항로짓 모형은 기존에 널리 사용되고 있기 때문에 본 연구에서는 별도로 기술하지 않으나, 자세한 내용은 Ben-Akiva and Lerman(1985)와 Train(2009)를 참고하면 된다. 특히 국내 통행 수단선택에 다항로짓 모형을 적용한 최근 연구는 Kim et al.(2018)에 요약되어 있다. 선행연구에서 기술한 것처럼 잠재계층모형은 선택행태에 있어 이질성을 고려하기 위해 다항로짓 모형을 확장한 것이다. 즉, 모집단이 모두 동일한 효용함수를 갖기 보다는(다항로짓 모형의 가정) 모집단 내에 선택의 효용함수가 다른 잠재계층이 존재한다고 가정하는 것이다.

잠재계층모형은 membership 모형과 class-specific choice 모형의 두 가지 부분으로 구성된다. 전자는 개인이 각 잠재계층에 속할 확률을 모형화하는 것이고, 후자는 각 잠재계층에서 개인이 각 대안을 선택할 확률을 모형화하는 것이다. 즉, 기존의 다항로짓 모형에서 모든 개인이 동일한 선택 효용함수(같은 파라미터)를 가정하는 반면, 잠재계층모형은 각 계층이 서로 다른 효용함수를 가질 수 있음을 가정한다(즉, 이질성을 의미한다). 본 연구에 적용된 잠재계층모형의 수단 선택 확률은 다음과 같이 표현된다.

$$P_n(i) = \sum_{k=1}^k P_{nk} \times P_n(i|k)$$

여기서, n 은 개인, i 는 통행수단, k 는 잠재계층($k = 1, 2, \dots, k$), P_{nk} 는 개인 n 이 잠재계층 k 에 포함될 확률, $P_n(i|k)$ 는 개인 n 이 잠재계층 k 내에서 수단 i 를 선택하는 조건부 확률이다. 각 P_{nk} 와 $P_n(i|k)$ 는 다음과 같이 표현된다.

$$P_{nk} = \frac{e^{\gamma'_k z_n}}{\sum_l e^{\gamma'_l z_n}}, \quad P_n(i|k) = \frac{e^{\beta'_{ki} x_n}}{\sum_j e^{\beta'_{kj} x_n}}$$

여기서 z_n 은 membership 모형의 설명변수 벡터, γ'_k 는 membership 모형의 파라미터 벡터, x_n 은 수단선택 모형의 설명변수 벡터, β'_{ki} 는 class-specific choice 모형의 파라미터 벡터이다. 대안 선택과 관련된 요소들이 membership 모형에 포함될지 class-specific choice 모형에 포함될지는 분석자가 결정해야 하며, 이는 연구의 합리적인 가설에 의해 결정된다. 본 연구에서는 목적지의 성격이 수단선택의 잠재계층을 결정한다고 가정하였기 때문에 z_n 은 목적지의 주거, 상업, 업무, 교육, 산업 시설물 면적 등의 토지이용 특성을 반영하였다. 또한 x_n 은 수단의 효용에 영향을 미치는 요소들(예를 들면, 수단 관련 변수 및 개인 속성)을 포함한다. 따라서 목적지의 성격이 잠재계층을 구분하며, 궁극적으로 계층간 다양한 요소들이 수단선택(파라미터)에 영향을 미치기 때문에 이질성을 반영하게 되는 것이다.

IV. 잠재계층모형 개발

1. 모형 선택

잠재계층모형에서 가장 중요한 부분 중 하나는 잠재계층의 개수를 결정하는 것이다. 모집단(population)에서 수단선택에 대한 잠재계층이 실제로 얼마나 존재하는지 알기 어렵기 때문이다. 또한 이론적으로 더 많은 잠재계층을 가정함에 따라 모형의 적합도가 지속적으로 증가하나, 잠재계층 개수가 증가함에 따라 모형의 복잡도(예, 파라미터의 개수)가 증가하기 때문에 모형으로서의 의미가 감소한다. 일반적으로 잠재계층의 개수를 정함에 있어 정량적으로는 정보지표(information criteria)가 고려되며, 정성적으로는 모형 파라미터들의 유의성과 해석가능여부가 고려된다. <Table 3>은 잠재계층의 개수에 따른 모형 지표들이다. 정량적 지표 측면에서는 모형의 Log-likelihood가 클수록, Akaike Information Criterion(AIC), Bayesian Information Criterion(BIC), Consistent AIC(CAIC)가 작을수록 모형이 더 선호된다고 볼 수 있다. 잠재계층이 증가함에 따라 Log-likelihood가 지속적으로

로 증가하고 AIC가 지속적으로 감소함을 알 수 있다. 즉, 모형이 정량적 지표로는 계속 개선된다는 것을 의미한다. 하지만, 기존연구에서 많이 이용된 모형의 복잡도에 대한 패널티를 더 고려하는 지표들(BIC 및 CAIC)에 따르면 두개의 잠재계층을 고려할 때의 모형이 가장 적절한 것으로 나타났다. 또한 모형의 계수들이 유의미하고 해석력이 좋아 본 연구에서는 두개의 잠재계층을 고려하였다.

<Table 3> Model indices by number of classes

Number of classes	Number of parameters	Log-likelihood	AIC	BIC	CAIC
1	22	-68257.37	136558.75	136760.99	136782.99
2	50	-67082.06	134264.13	134723.76	134773.76
3	78	-66960.93	134077.86	134794.88	134872.88
4	106	-66832.52	133877.05	134851.47	134957.47

2. 모형 추정 결과

<Table 4>와 <Table 5>는 잠재계층모형을 추정한 결과이다. 일반적으로 그룹을 미리 나누고 모형을 추정하는 것과 달리 잠재계층모형은 membership 모형(Table 4)와 class-specific choice 모형(Table 5)가 동시에 추정된다. 두 개의 잠재계층은 각각 표본의 68%, 32%를 차지했으며, 수단선택에 있어 Class 1의 자가용과 대중교통의 부담비가 대략 절반씩(46:54)인 반면, Class 2의 부담비는 20:80으로 대중교통의 비중이 더욱 높았다. 이 두가지의 잠재계층을 구분하는 membership 모형의 주요 설명변수로써 5가지의 목적지 시설물 관련 변수가 포함되었다. 이 중 하나의 변수(education facility)를 제외한 나머지의 변수들은 모두 $\alpha = 0.01$ 의 수준에서 통계적으로 유의했다. 파라미터의 방향성을 고려할 때 목적지에 업무시설, 상업시설의 총면적이 넓을수록, 주거시설의 총면적이 적을수록 Class 2에 속할 확률이 높아졌다. 즉, 목적지의 특성에 따라 수단선택에 대한 잠재계층이 구분될 수 있음을 암시한다.

<Table 4> Estimation results of class membership model (Reference class: Class1)

Category		Class1		Class2	
Class size		68.0%		32.0%	
Mode share	Auto	46.0%		20.4%	
	Bus	23.1%		17.3%	
	Metro	30.9%		62.2%	
		Coefficient	z-value	Coefficient	z-value
Intercept		-	-	2.915**	4.11
Log(Residential facility)		-	-	-0.974*	-14.32
Log(Commercial facility)		-	-	0.230**	4.19
Log(Business facility)		-	-	0.603**	13.39
Log(Education facility)		-	-	-0.028	-1.13
Log(Industry facility)		-	-	-0.059**	-4.78

Note: ** p-value<0.01, * p-value<0.05. Five membership variables implicitly mean the logarithm of facility area at the destination zone by facility use. To avoid negative infinite for zero values (i.e. log(0)=negative infinite), small number (+1) is added to original values.

<Table 5> Estimation results of class-specific choice models(Reference mode: Auto)

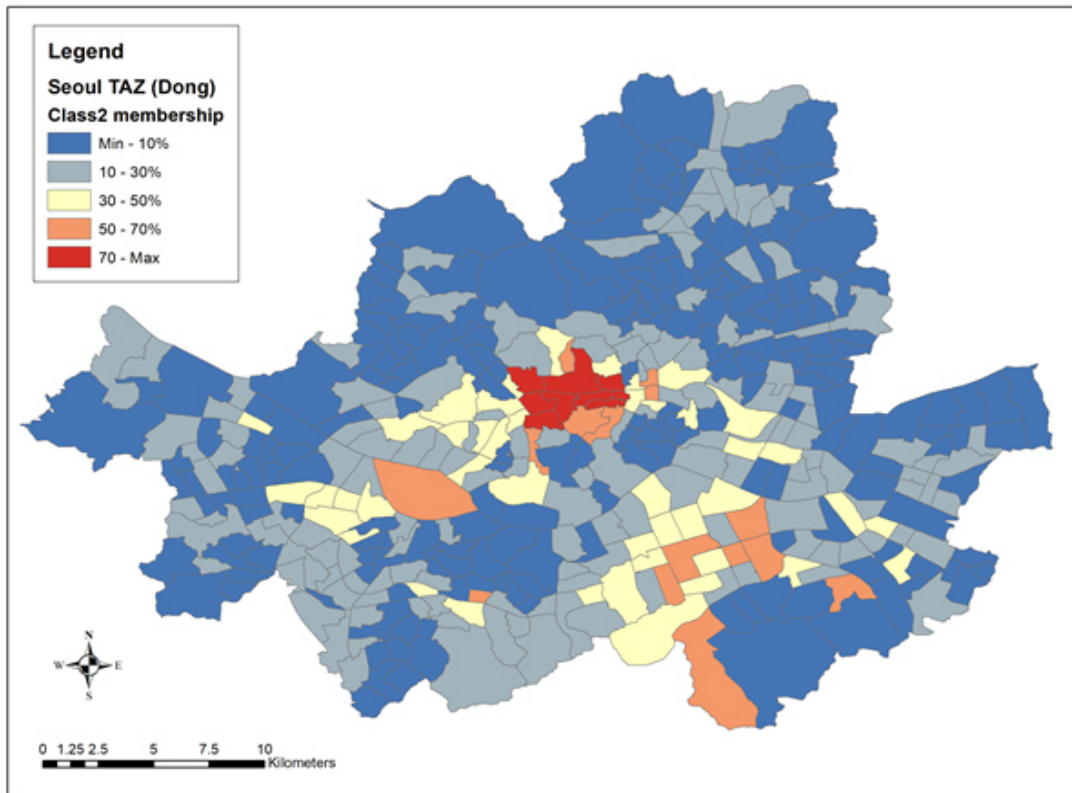
Variable	Category	Choice	Class 1		Class 2	
			Coefficient	z-value	Coefficient	z-value
Intercept	Intercept	Bus	3.900**	47.31	3.881**	13.9
		Metro	4.076**	50.02	4.982**	18.22
Car ownership	Own car=1	Bus	-2.549**	-42.38	-2.407**	-11.04
		Metro	-2.573**	-42.85	-2.271**	-10.77
Metro accessibility (reference: 10min or less)	11-15 min	Bus	0.137**	3.12	0.144	1.41
		Metro	-0.400**	-8.78	-0.171*	-2.08
	16 min or more	Bus	0.256**	6.05	0.307**	2.97
		Metro	-0.728**	-14.77	-0.293**	-3.38
Gender	Male=1	Bus	-1.566**	-45.39	-1.165**	-12.49
		Metro	-0.909**	-26.78	-0.809**	-10.38
Income (reference: below 2 million won)	2-5 million won	Bus	-0.365**	-8.96	-0.132	-1.16
		Metro	-0.264**	-6.42	-0.08	-0.83
	5 million + won	Bus	-1.004**	-17.64	-0.658**	-5.12
		Metro	-0.606**	-11.63	-0.716**	-6.77
Age cohort (reference: 19-29)	Age (30-49)	Bus	-1.289**	-25.12	-0.977**	-6.77
		Metro	-1.537**	-31.75	-1.055**	-8.32
	Age (50-64)	Bus	-1.077**	-19.17	-1.019**	-6.53
		Metro	-1.569**	-29.36	-1.392**	-10.32
	Age (65+)	Bus	-0.514**	-5.23	-0.723**	-2.69
		Metro	-0.662**	-7.2	-0.647**	-2.9
Occupation	White collar=1	Bus	0.097**	2.67	0.049	0.62
		Metro	0.526**	15.93	0.305**	4.86
Model summary						
Number of cases		72,595				
Log-likelihood		-67082.1				
ρ^2		0.16				

Note: **p-value<0.01, *p-value<0.05.

Class-specific choice 모형들을 살펴볼 때 대부분의 파라미터들이 통계적으로 유의했으며, 부호의 방향성도 합리적이었다. 전반적으로 차량을 소지하면, 집에서 지하철역까지의 거리가 증가할 수록, 남자이며, 소득이 증가할 수록 자가용을 선택할 확률이 증가하였다. 이는 기존의 연구결과와 일관된 것으로 보인다. 모형 추정 결과에서 나타난 것과 같이 각 잠재계층은 수단선택에 대한 다른 효용함수를 가지고 있음을 알 수 있다. 즉, 기존의 단일모형에서는 각 요소가 수단선택의 효용에 미치는 영향을 모집단의 단일 평균으로 추정하는 것과 달리 잠재계층모형에서는 각 잠재계층의 효용함수를 별도로 추정할 수 있다. 따라서 수단선택 함수의 이질성을 고려함에 따라 더 현실적인 모형이 될 가능성이 높다.

Class-specific choice 모형들을 비교해볼 때 특정한 패턴을 찾을 수 있는데, Class 1이 Class 2보다 전반적으로 더 영향도가 높은 파라미터들을 가지고 있음을 확인할 수 있다. 즉, Class 1이 Class 2에 비해 수단에 관련된 요소(차량소유여부, 지하철역 접근성)와 사회경제적 특성(성별, 소득, 연령, 직업)의 변화에 더 민감하다고

볼 수 있다. 이는 잠재계층의 특성과 연결지어 생각해 볼 필요가 있다. <Fig. 2>는 Class 2의 소속확률을 서울 지도에 표현한 것이다. 즉, 목적지 존의 색이 붉은 색일수록 Class 2에 속할 확률이 높고 파란 색일수록 Class 1에 속할 확률이 높다. 잠재계층을 구분짓는 목적지의 분포 중 눈에 띄는 패턴은 목적지가 도심(종로, 강남, 여의도 등)인 경우 Class 2에 속할 확률이 높다는 것이다. 특히 목적지가 종로와 명동일대의 경우 Class 2에 속할 확률이 70% 이상이다.



<Fig. 2> Geographical distribution of class 2 membership probability

각 잠재계층이 어떤 사람들로 구성되고 어떤 특징을 가지고 있는지에 대해 이해하는 것은 잠재계층모형의 핵심중 하나이다. 이를 위해 잠재계층별 특징들에 대한 기댓값을 추정하고 비교하는데, 각 개인이 각 잠재계층에 속할 소속확률을 활용하여 확률가중평균으로 기댓값을 계산한다. 소속확률은 두가지(prior and posterior probabilities)로 분류할 수 있는데, 본 연구는 Bhat(1997)가 제안한 prior probability를 적용한 가중평균 공식을 활용하였다(Table 6).

$$\bar{x}_k = \frac{\sum_n p_{nk} x_n}{\sum_n p_{nk}}$$

<Table 6> Probability-weighted class-specific mean/share

Category		Pooled sample (sample mean)	Latent class model	
			Class1	Class2
Cluster size		100.0%	68.0%	32.0%
Mode share	Auto	37.8%	46.0%	20.4%
	Bus	21.2%	23.1%	17.3%
	Metro	40.9%	30.9%	62.2%
Residential facility of destination (square meters)		516,271.5	597,128.9	379,165.2
Commercial facility of destination (square meters)		267,294.5	208,348	453,448.9
Business facility of destination (square meters)		51,605.9	22,225.9	308,148.4
Educational facility of destination (square meters)		37,682.6	35,966.1	41,600.4
Industrial facility of destination (square meters)		1,616.8	1,258.0	2,752.6
Car ownership (yes=1)		83.9%	84.0%	83.7%
Metro accessibility (10min or less)		70.4%	69.9%	71.5%
Metro accessibility (11-15min)		15.1%	15.1%	14.9%
Metro accessibility (16min or more)		14.6%	15.0%	13.6%
Gender (male=1)		65.8%	65.4%	66.7%
Income (Below 2 million won)		21.0%	21.8%	19.3%
Income (2-5 million won)		59.5%	59.6%	59.3%
Income (5 million won +)		19.5%	18.5%	21.5%
Age (19-29)		16.5%	16.1%	17.6%
Age (30-49)		55.5%	54.9%	56.6%
Age (50-64)		24.7%	25.6%	22.6%
Age (65+)		3.3%	3.4%	3.2%
Occupation (White collar =1)		36.5%	34.2%	41.5%

목적지의 특성을 살펴보면, Class 1의 평균 주거시설면적은 597,128.9m²로 Class 2의 평균면적(379,165.2m²)의 1.6 배 수준이다. 반대로 Class 2는 평균적으로 더 넓은 상업, 업무, 교육, 산업 시설면적을 가지고 있다. 이는 <Fig. 2>의 결과와 일관된 것으로 보인다. 수단선택의 설명변수를 살펴보면, Class 2에 속하는 사람들이 평균적으로 더 지하철역의 접근성이 좋고, 남자이며, 더 고소득이며, 사무직 직종이다. 하지만 직업을 제외한 나머지 변수의 경우 그 차이가 2% 이내로 크다고 보기는 어렵다. 이러한 잠재계층들의 특성을 고려해볼 때, 평균적으로 도심으로 출근하는 사람들(Class 2)은 수단선택에 미치는 요소들에 대해 덜 민감하다고 볼 수 있다. 다시 말하면, 선택에 미치는 요소들에 영향을 많이 받기보다는 대체로 대중교통을 이용하는 것이다. 이는 여러 이유가 존재할 수 있지만, 도심으로 출근하는 경우 차량을 이용할 때의 비효용(예, 주차비용, 주차부담, 혼잡으로 인한 불편)이 높아 다른 외부적인 요인에 덜 영향을 받고 대중교통을 주로 이용하는 것으로 보인다. 모형을 종합적으로 볼 때 유사한 사람이 같은 출발지에서 출근함에도 불구하고 목적지에 따라 수단선택이 다를 수 있다는 것을 암시한다. 이는 수단선택의 이질성을 잘 나타내고 있는 것으로 평가된다.

V. 결 론

본 연구에서는 서울시내 출근통행에 있어 수단선택의 이질성에 대하여 분석하였다. 특히 선행연구를 통해 목적지의 성격과 수단선택에 대한 연구와 통행행태의 이질성에 대한 연구가 미비했음을 확인하였다. 이에 따라 목적지 시설물 면적에 대한 변수들을 활용하여 수단선택에 대한 잠재계층을 구분하고, 각 잠재계층이 어떻게 수단선택을 하는지를 잠재계층별 효용함수를 통해 살펴보았다. 이를 위하여 확률적 모형인 잠재계층 모형을 적용하였고, 유의미한 다른 특성의 잠재계층이 존재함을 확인하였다. 여러 정량적, 정성적 평가를 통해 두개의 잠재계층을 정의하였다. Class 1은 승용차와 대중교통의 수단분담이 비슷하였고, 목적지가 상대적으로 도심이 아닌 주거시설이 많고 업무/상업 시설이 적은 지역인 경우였고, class 2는 주로 대중교통의 분담률이 높고, 목적지가 상대적으로 더 도심(업무/상업 시설이 많은 지역)인 경우였다. 특히 지도를 통해 목적지가 주요 도심(종로, 강남, 여의도)인 경우 class 2에 속할 확률이 매우 높음을 확인하였다. 두 계층 모두 수단과 관련된 변수와 개인의 특성이 수단선택에 유의미한 영향을 미쳤으며, 그 영향도가 다름을 확인하였다. 특히, class 2는 class 1에 비하여 상대적으로 여러 요소들에 덜 민감한 것으로 나타났다. 이는 결과적으로 수단선택의 이질성을 나타내며, 유사한 사람이 같은 출발지에서 출근함에 불구하고 목적지에 따라 수단선택이 다를 수 있다는 것을 암시한다. 이러한 수단선택의 이질성은 교통정책(가령 대중교통 공급, 자가용 이용 억제 정책)과 사회구성층의 변화(고령화)에 따라 수단분담률의 변화가 지역(통행의 목적지)에 따라 다르게 나타날 수 있음을 시사한다. 따라서 교통정책의 효과 평가나 미래의 통행행태의 변화에 대한 예측에 있어 이질성에 대한 고려가 필요할 것으로 보인다.

모형측면에서는 잠재계층을 구분함에 따라 기존의 단일모형에 비해 적합도가 높고 통행행태에 대한 설명력이 더 우수했음을 확인하였다. 비확률적 분류법에 비해 잠재계층모형은 사전에 미리 계층을 임의로 구분할 필요가 없고, 확률개념을 도입함으로써 더욱 합리적인 모형구축이 가능하다. 물론 잠재계층을 구분하는 변수를 결정하고 잠재계층의 개수를 정하는 것이 다소 주관적이거나, 수단선택에 대한 의사결정이 다른 그룹을 실제로 연구자가 알 수 없다는 측면을 고려한다면, 이 모형이 통행행태측면에서 잠재계층을 발견하는데 도움이 됨을 알 수 있었다.

본 연구에서는 연구 목적을 위해 분석 대상을 한정하였으나 이 또한 연구의 한계점으로 생각되며 이를 보완하고 해결하기 위한 연구가 추가적으로 필요할 것으로 생각된다. 첫 째로 본 연구는 통행의 주수단에 초점을 맞춰 세 가지 통행수단에 대해 분석하였다. 하지만, 서울시의 복잡한 대중교통 네트워크를 고려할 때 실질적으로 두개 이상의 수단(예, 버스-지하철, 버스-버스, 자가용-지하철)이 복합적으로 이용되는 경우가 많다. 따라서 수단선택에 있어 환승을 고려한 통행사슬(trip chain)을 고려한다면, 보다 현실적인 결과를 도출하고 다양한 통행행태의 잠재계층을 발견할 것으로 생각된다. 또한 본 연구는 출근통행에 한정하였으나, 연구모형을 다양한 여가통행에도 적용 가능할 것이다. 특히 출근통행은 고정적 시간/장소와 의무성의 성격이 강하나 여가통행의 경우 개인의 의사결정이 더 결부될 것으로 판단된다. 따라서 여가통행의 수단선택에는 더 다양한 잠재계층이 존재할 가능성이 있다. 추가적으로 모형의 변수에 있어 앞서 언급한 바와 같이, 선택되지 않은 대안들의 통행시간과 통행비용에 대한 정보가 없어 이들 변수를 본 모형에서 제외하였다. 향후 연구에서 통행시간 및 통행비용을 모형에 반영한다면, 모형의 적합도 및 예측력 개선은 물론 잠재계층별 통행시간 및 통행비용의 민감도 분석도 가능할 것으로 판단된다. 마지막으로 보다 최근에 완료된 2016년 가구통행실태조사에 대한 분석과 본 연구를 연계한다면, 서울 통근행태의 전반적인 변화추세와 잠재계층의 변화를 확인해 볼 수 있을 것으로 생각된다. 특히 토지이용의 변화나 대중교통시스템의 변화(가령9호선의개통)로 인한 특정 목적지에 대한 통행행태의 변화를 살펴 볼 수 있을 것으로 기대한다.

REFERENCES

- Anowar S., Yasmin S., Eluru N. and Miranda-Moreno L. F.(2014), “Analyzing car ownership in Quebec City: a comparison of traditional and latent class ordered and unordered models,” *Transportation*, vol. 41, no. 5, pp.1013-1039.
- Bae Y. G., Jeong J. H. and Kim H. J.(2010), “Latent Class Analysis for Mode Choice Behavior,” *Journal of Korean Society of Transportation*, vol. 28, no. 3, pp.99-107.
- Bae Y. G., Jeong J. H. and Kim H. J.(2011), “The Influence of built environment on Mode Choice: Latent Class Approach,” *Journal of the Korean Regional Science Association*, vol. 27, no. 2, pp.99-111.
- Ben-Akiva M. and Lerman S. R.(1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press.
- Ben-Akiva M., McFadden D., Train K., Walker J., Bhat C., Bierlaire M. and Daly A.(2002), “Hybrid choice models: Progress and challenges,” *Marketing Letters*, vol. 13, no. 3, pp.163-175.
- Bhat C. R.(1997), “An endogenous segmentation mode choice model with an application to intercity travel,” *Transportation Science*, vol. 31, no. 1, pp.34-48.
- Carrasco J. A. and de Dios Ortúzar J.(2002), “Review and assessment of the nested logit model,” *Transport Reviews*, vol. 22, no. 2, pp.197-218.
- Cervero R. and Kockelman K.(1997), “Travel demand and the 3Ds: density, diversity, and design,” *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, pp.199-219.
- Choo S. and Mokhtarian P. L.(2004), “What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice,” *Transportation Research Part A: Policy and Practice*, vol. 38, no. 3, pp.201-222.
- Ewing R. and Cervero R.(2010), “Travel and the built environment: A meta-analysis,” *Journal of the American Planning Association*, vol. 76, no. 3, pp.265-294.
- Greene W. H. and Hensher D. A.(2003), “A latent class model for discrete choice analysis: Contrasts with mixed logit,” *Transportation Research Part B: Methodological*, vol. 37, no. 8, pp.681-698.
- Handy S., Cao X. and Mokhtarian P.(2005), “Correlation or causality between the built environment and travel behavior? Evidence from Northern California,” *Transportation Research Part D: Transport and Environment*, vol. 10, no. 6, pp.427-444.
- Hensher D. A. and Greene W. H.(2003), “The mixed logit model: the state of practice,” *Transportation*, vol. 30, no. 2, pp.133-176.
- Johansson M. V., Heldt T. and Johansson P.(2006), “The effects of attitudes and personality traits on mode choice,” *Transportation Research Part A: Policy and Practice*, vol. 40, no. 6, pp.507-525.
- Kim S. H. and Mokhtarian P. L.(2018), “Taste heterogeneity as an alternative form of endogeneity bias: Investigating the attitude-moderated effects of built environment and socio-demographics on vehicle ownership using latent class modeling,” *Transportation Research Part A: Policy and Practice*, vol. 116, pp.130-150.
- Kim S., Choo S., Kim J. and Han J.(2018), “An Analyzing Factors to Affect Trip Mode Chaining Behavior Using Travel Diary Survey Data in Seoul,” *Journal of the Korea Institute of Intelligent Transport Systems*, vol. 17, no. 1, pp.55-70.

- Lanzini P. and Khan S. A.(2017), “Shedding light on the psychological and behavioral determinants of travel mode choice: A meta-analysis,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 48, pp.13-27.
- Popovich N. D. and Handy S.(2015), “Downtown, strip centers, and big-box stores: Mode choice by shopping destination type in Davis, California,” *Journal of Transport and Land Use*, vol. 8, no. 2, pp.149-170.
- Reilly M. and Landis J.(2003), *The influence of built-form and land use on mode choice: Evidence from the 1996 Bay Area Travel Survey*, Institute of Urban and Regional Development at UC Berkeley.
- Sun B., Ermagun A. and Dan B.(2017), “Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai,” *Transportation Research Part D: Transport and Environment*, vol. 52, pp.441-453.
- The Korea Transport Institute(2011), *2010 Interregional Passenger OD Travel Survey*, available at <https://www.ktdb.go.kr/www/contents.do?key=60>.
- The Korea Transport Institute(2012), *2011 National Transport Demand Survey and Database Construction Project*, available at <https://www.ktdb.go.kr/www/contents.do?key=24>.
- Train K. E.(2009), *Discrete Choice Methods with Simulation*, Cambridge University Press.
- Vij A. and Walker J. L.(2014), “Preference endogeneity in discrete choice models,” *Transportation Research Part B: Methodological*, vol. 64, pp.90-105.
- Vij A. and Walker J. L.(2016), “How, when and why integrated choice and latent variable models are latently useful,” *Transportation Research Part B: Methodological*, vol. 90, pp.192-217.
- Vij A., Carrel A. and Walker J. L.(2013), “Incorporating the influence of latent modal preferences on travel mode choice behavior,” *Transportation Research Part A: Policy and Practice*, vol. 54, pp.164-178.
- Walker J. L. and Li J.(2007), “Latent lifestyle preferences and household location decisions,” *Journal of Geographical Systems*, vol. 9, no. 1, pp.77-101.
- Zhang M.(2004), “The role of land use in travel mode choice: Evidence from Boston and Hong Kong,” *Journal of the American Planning Association*, vol. 70, no. 3, pp.344-360.