

■ 論 文 ■

부산광역권 교통수단선택모형의 정립과 모수추정에 관한 연구

Model Specification and Estimation Method for Traveler's Mode Choice Behavior
in Pusan Metropolitan Area

김 익 기

(한양대학교 교통공학과 교수)

김 강 수

(교통개발연구원 연구위원)

김 형 철

(한양대학교 교통공학과 석사과정)

목 차

I. 서론	1. 모형 1 : 기본모형
II. 자료	2. 모형 2 : 개선모형 -1
1. 개요	3. 모형 3 : 개선모형 -2
2. 자료 검증 및 주 교통수단의 파악	4. 모형 4 : 네스티드 로짓모형
3. 자료 보완	V. 장래 교통수단 분담율의 예측
III. 모형정립	VI. 결론
IV. 모형추정결과	참고문헌

Key Words : 로짓모형, 개별행태자료, 네스티드 로짓모형, 가구통행실태조사, 수단선택모형

요 약

수단선택모형은 교통 분석 및 수요예측을 위해 필수적 모형이나 그 추정과정이나 실제적인 방법론에 대한 깊이 있는 논의가 우리나라에서는 부족하였다. 1980년대부터 로짓모형을 이용한 수단선택모형이 광범위하게 이용되고 있지만 기본적인 기초 조사자료 구축에서부터 수단선택 모형설정 및 정산까지 추정과정에 대한 투명하고 상세한 논의가 부족하였다. 그 결과 구축된 교통수단선택모형에 대한 신뢰성이 제기되거나 혹은 구축된 결과를 적용하여 분석에 이용하는데 있어 오류가 발생하는 등 적잖은 부작용도 있었다.

본 논문의 목적은 부산광역권의 통행행태 개별자료를 이용하여 교통수단선택 모형의 정산 및 구축에 대한 상세하고 투명한 계시를 통해 부산광역권 교통정책에 실질적인 적용이 가능한 예측모형을 제시하는 데 있다. 즉 수단선택모형 구축을 위한 조사방법, 변수설정, 모형정산 및 검증 그리고 예측 등의 일련의 과정을 투명하고 명확하게 제시하고, 정책분석에 활용 가능한 모형을 제시함으로써 교통수단선택모형의 구축 방법론 발전에 초석이 되고자 하는 것이 본 연구의 주요 목적이다. 이를 위해, 선택되지 않은 교통수단에 대한 자료가 일반적으로 조사 자료에는 없다는 점을 보완하기 위한 비선택대안의 속성 추정방법을 이 논문에서 제안하였다.

본 논문에서는 1999년 및 2001년 부산울산권의 가구통행실태조사 자료를 활용하여 네스티드 모형구조를 포함하여 총 4개의 모형이 단계별로 구축되었으며 또한 각각 통계적으로 검증되어졌다. 또한 제시된 예측모형의 현실성을 검증하기 위해 효용함수에 의한 시간가치를 계산하여 기존 연구의 시간가치 규모와 비교하였으며, 또한 분류(classification) 방법에 의한 예측 기법을 이용하여 조사년도의 통행행태를 추정하고 조사 자료와 비교하여 모형의 예측력에 대해서도 검증을 하였다.

Mode choice Analysis is essential analysis stage in transportation demand forecasting process. Therefore, methods for calibration and forecasting of mode choice model in aspect of practical view need to be discussed in depth. Since 1980s, choice models, especially Logit model, are spread widely and rapidly over academic area, research institutes and consulting firms in Korea like other developed countries in the world. However, the process of calibration and parameter estimation for practical application was not clearly explained in previous papers and reports. This study tried to explain clearly the calibration process of mode choice step by step and suggested a forecasting mode choice model that can be applicable in real policy analysis by using household survey data of Pusan metropolitan area. The study also suggested a way of estimating attributes which was not observed during the household survey commonly, such as travel time and cost of unchosen alternative modes. The study summarized the statistical results of model specification for four different Logit models as a process to upgrade model capability of explanation for real traveler's choice behaviors. By using the analysis results, it also calculated the value of travel time and compared them with the values of other previous studies to test reliability of the estimated model.

1. 서론

수단선택모형은 교통 분석 및 수요예측을 위해 필수적 모형이나 그 추정과정이나 실제적인 방법론에 대한 깊이가 있는 논의가 우리나라에서는 부족하였다. 1980년대부터 로짓모형을 이용한 수단선택모형이 광범위하게 이용되고 있지만 기본적인 기초 조사자료 구축에서부터 수단선택 모형설정 및 정산까지 추정과정에 대한 투명하고 상세한 논의가 우리나라에서는 부족하였다. 그 결과 정산 구축된 교통수단선택모형에 대한 신뢰성이 제기되거나 혹은 구축된 결과를 적용하여 분석에 이용하는데 있어 오류가 발생하는 등 적잖은 부작용도 있었다. 따라서 본 논문의 목적은 부산-울산 광역권의 통행행태 개별자료를 이용하여 로짓모형 형태의 교통수단선택 모형의 정산 및 구축에 대한 상세하고 투명하게 제사를 통해, 부산과 울산광역시 교통정책에 실질적인 적용이 가능한 예측모형을 제시하는데 있다. 또한 수단선택모형 구축을 위한 조사방법, 변수설정, 모형정산 및 검증, 그리고 예측 등의 일련의 과정을 투명하고 명확하게 제시하여, 정책분석에 활용 가능한 모형을 제시함으로써 교통수단선택모형의 구축 방법론 발전을 위한 활발한 논의의 계기를 제공하는 것도 본 연구의 주요 목적이다. 이와 같이 구축된 교통수단선택 예측모형은 도시 및 지역교통의 시설물 계획 및 운영정책 분석을 위한 교통수요분석에 응용될 수 있을 것이며 또한 교통수단 분담을 예측을 통한 수단별 교통정책 분석에도 활용될 수 있기를 기대한다.

교통수단선택 모형과 관련된 우리나라에서의 연구로 윤대식(1997)은 네스티드 로짓모형을 이용한 통근통행자의 수단선택모형을 구축하였으며, 조중래·김재만(1998)은 서울과 일산 신도시의 출근통행에 대하여 다항로짓모형을 이용하여 지역간 이전 가능성을 분석하였으며, 김대용의 다수(1999)는 대구광역시외 취업자를 대상으로 모형의 이전 가능성을 검토하였다. 하지만 이와 같은 대부분의 기존 연구에서는 주로 로짓모형이 갖고 있는 "비관련대안으로부터의 독립성" (IIA 속성)으로 인한 현실 정책적용의 한계성을 극복하기 위한 방안을 제시하거나 조사 자료를 통한 효용함수 구축과 결과 해석 그리고 모형 전이성 등 정책분석의 응용에 초점을 둔 연구가 대부분이었다. 즉 기존 연구에서는 대부분 분석 결과에 초점을 두고 있어 기초자료 구축과정과 모형정산 과정에 대한 상세한 설명이 생략됨에 따라 적용된

자료의 신뢰성과 구축된 교통수단선택 모형 그 자체에 대한 신뢰성을 논의하기가 힘들다. 이와 같이 입력자료의 구체적 정의와 모형 정산의 단계별 상세한 논의가 없게 되면 수단선택모형구축을 위한 조사방법, 변수설정, 모형정산, 예측기법, 입력자료 예측 등의 일련의 분석 과정에 임의적 방법론과 결과 해석이라는 여지가 남게 되어 모형의 신뢰성에 대한 논란이 있을 수 있게 된다. 따라서 본 연구에서는 가능한 자료의 구축에서 모형의 정산 기법 및 예측방법에 대해 명료하고 구체적으로 각 분석과정을 논의하였다.

본 연구에서 사용된 자료의 형태는 존 기반의 집합자료 (zone-based aggregate data)가 아닌 개인통행별 자료인 개별자료(disaggregate data)가 사용되어 통계 이론적으로 우수한 모형을 구축하고자 하였으며, 예측 시에는 집합화 오차(aggregation error)를 줄이는 방법도 고려하였다. 모형 정산 시에 입력 자료로써 그동안 언급되지 않았던 선택되지 않은 대안에 대한 속성 자료를 추정하여 구축하는 과정을 상세하게 제안하였다. 또한 교통수단선택 모형 안에 포함되는 독립변수에 대해서도 그 값이 어떻게 추정되어 모형정산 과정에 적용하였는가를 투명하게 설명함으로써 모형정산 과정에서의 논란 여지를 가능한 최소화 하였다. 또한 교통수단선택 모형으로 적용된 로짓모형에 있어 효용함수 형태와 변수 설정(model specification) 과정에 있어서도 통행자의 논리적인 교통수단선택 행태원리를 기반으로 기본모형을 설정하였고, 이 기본모형을 기초로 더 좋은 모형을 찾아가는 방법과 과정을 명확하게 제시하여 본 논문에서 최종적으로 제안된 예측모형의 적합성을 보여주었다. 즉 독립변수설정 및 네스티드 로짓모형을 포함한 모형구조 등을 다양하게 변형하여 모형 개발을 시도하고 대안 예측모형 간의 통계적 우수성을 비교하여 조사자료를 가장 잘 설명하는 예측모형을 제안하고자 한 것이다. 또한 본 연구에서 제안한 예측모형의 현실성을 검증하기 위해 통행시간변수에 대한 계수값과 통행비용변수에 대한 계수값을 이용하여 시간가치 값을 구한 후 기존 연구의 결과와 비교하여 적합성을 검토하였다. 그리고 모형의 예측력을 검증하기 위해 분류(classification) 방법에 의한 예측 기법을 이용하여 조사년도의 통행 행태를 추정하고 조사 자료와 비교하여 모형의 예측력에 대해서도 검증을 하였다. 이와 같은 입력자료, 모형정산, 분석결과 해석 그리고 예측 방법 등에 대해 투명하게 설명함으로써 이 논문에서 제

시한 모형보다 더욱 우수한 모형을 향후에 개발하는데 시작점이 될 수 있도록 충분한 분석 정보를 제공하였다. 이와 같은 일련의 과정이 교통전문가 간에 활발한 논의가 될 수 있는 동기를 본 연구에서는 제공할 수 있기를 기대한다.

II. 자료

1. 개요

본 연구에서 사용된 자료는 교통개발연구원의 국가교통DB센터에서 조사한 1999년 및 2001년 부산·울산광역시 및 인접 시·군에 대한 가구통행실태조사 자료이며 (교통개발연구원(2000, 2002)), 본 논문에서는 이 지역은 간단히 부산권이라 명칭하였다. 좀 더 구체적으로는 부산, 울산, 경주, 창원, 마산, 진해, 김해, 양산시 등 총 8개의 시가 조사되었으며 행정동과 면을 기준으로 구축되어, 외부준 포함 총 406개 존으로 구분되어 있다. 부산권과 울산권을 함께 분석한 이유는 국가연구기관의 자료 분류와 분석이 부산권과 울산권을 한 권역으로 묶어서 분석하고 있으므로 본 연구에서도 일관성을 유지하기 위해 동일한 권역으로 분석하였으며, 또한 개별자료 분석의 특성 상 각 개별의 교통 환경적 요인을 독립변수 및 대안집합에서 반영될 수 있다고 판단하여 하나의 권역으로 분석하는데 큰 무리가 따

〈표 1〉 가구통행실태조사자료 중 교통수단 선택모형 정산을 위해 필요한 자료의 내용

자료	내용
가구자료	○ 소득관련 자료 : 연소득, 주택종류 및 형태, 주거 면적, 가정의 직업 등 ○ 차량보유대수 등
개인자료	○ 나이, 직업, 성별, 승용차 이용 가능성 여부, 운전 면허 소지여부 등
통행자료	○ 선택교통수단, 선택교통수단의 속성 (통행시간, 통행비용, 접근교통수단 및 접근통행시간), 비 선택 대안 교통수단의 집합, 비 선택 대안 교통수단의 속성 등
통행목적	○ 가정기반의 통행 - 가정기반 출퇴근통행, 가정기반 등하교통행, 가정기반 쇼핑통행, 가정기반 여가 통행(오락, 친교), 가정기반 기타통행 ○ 비 가정기반의 통행 - 비 가정기반 업무통행, 비 가정기반 쇼핑통행, 비 가정기반 여가통행, 비 가정기반 기타통행

르지 않는다고 생각하여서이다. 조사된 총 자료의 수는 130,811 자료이나 이후 모형구축을 위해 자료에 대한 검증 및 보완 과정을 거쳐 유효 자료만을 정리하여 분석이 이루어졌다. 최종적으로 수단선택모형 정산에 사용된 자료의 수는 9,236 자료이다. 가구통행실태 조사자료 표본의 일반적 특성 및 분포는 교통개발연구원(2000, 2002)에 제시되어 있다. 교통수단선택 모형구축에 필요한 내용을 가구 및 개인의 사회경제적 특성, 교통수단별 통행 서비스의 속성 및 통행목적으로 구분하여 정리한 내용은 〈표 1〉과 같다.

2. 자료 검증 및 주 교통수단의 파악

응답자의 응답 한계성, 조사비용의 한계, 설문지의 단순화 필요성 등에 의해 교통수단 선택모형 정산에 필요한 모든 자료를 조사하지 못하는 것이 일반적이다.

따라서 모형 정산을 위해서는 일부 자료의 변경 혹은 추정이 필요하다. 먼저, 유효자료를 획득하기 위한 자료의 검증이 필요하다. 본 연구에서는 응답자의 자료 입력과정에서의 오류, coding 작업상의 오류 및 논리적 검증을 통해서 불합리한 값 등에 대한 수정 작업이 수행되었다. 예를 들어 개별 통행자에 대해 선택된 교통수단의 통행시간의 산출결과 5시간을 초과하는 경우는 비정상적인 통행 혹은 오류에 의한 부적절한 값이라 판단하여 전체 통행시간이 5시간(300분)을 초과하는 통행은 분석 자료에서 제거하였다.

다음으로는 분석의 용이성을 위해 유사 교통수단들을 하나의 교통수단으로 통합하여 분석하였다. 이는 단일 목적통행 하에서 복합 교통수단을 이용한 경우의 모형구축이 현실적으로 어렵기 때문이다. 즉 주 교통수단(Line-haul mode)과 접근 교통수단(access mode)으로 구분하여 주 교통수단이 될 가능성이 높은 교통수단별로 우선순위를 정하여, 우선순위가 높은 교통수단이 주 교통수단이 되고, 나머지는 접근 교통수단으로 간주하여 분석하였다. 본 연구에서 사용한 주 교통수단 가능성이 높은 교통수단 우선순위는 미국 DOT(Department of Transport)(FHWA, 1977)에서 제안한 지하철, 버스, 택시, 승용차, 보행의 순서로 사용하였다. 예를 들면 버스-지하철-보행으로 출발지에서 도착지까지 교통수단을 이용하였으면 지하철이 주 교통수단이 되고, 버스와 보행이 접근수단이 되는 것이다.

3. 자료 보완

수단선택모형 구축을 위해서는 각 교통수단 선택을 설명하는 설명변수의 구축이 필요하다. 서울시정개발연구원 연구(2001) 및 일반적 보고서에서 승용차의 통행비용은 유류비, 유지 정비비, 주차비용, 통행료, 사고처리비용, 범칙금과 과태료, 시간비용 등과 같은 항목으로 구분하고 있으나 응답자가 인지하지 못해 응답할 수 없는 경우가 대부분이다. 따라서 본 연구에서는 승용차의 고정 및 유동비용에서 운전자가 통행별로 인지하는 비용(perceived cost)이 통행목적별로 통행자가 느끼는 시간가치와 비례하거나 유사한 값이 될 것으로 가정하였다. 즉 승용차 통행자는 특정 목적통행에 대해 통행으로 소요하는 시간의 시간가치만큼 기꺼이 지불할 의사(willingness to pay)가 있다고 가정하고 승용차 비용을 유추 추정하였다. 즉 승용차의 경우는 간접적 비용(indirect cost)의 부분이 고속도로 통행요금과 같은 직접 비용(direct cost)보다 일반적으로 크기 때문에 비용적 측면을 실감나게 느끼지 못하는 경우가 많다. 그러므로 본 연구에서는 통행목적에 따라 승용차를 이용함으로써 절약될 수 있는 통행시간의 시간가치에 통행자가 더 실질적으로 민감하게 영향을 받는다고 가정을 하고 분석하였다. 시간가치의 값에 대해서는 현재 보편적으로 활용되고 있는 한국개발연구원(1999)에서 제시한 업무통행 시간가치는 9,750원/시간, 비업무통행 시간가치는 업무통행가치의 1/3인 3,250원/시간을 활용하여 승용차 통행비용을 추정하였다¹⁾. 이와 같은 승용차 통행비용 추정방법은 같은 O/D 간이라도 통행목적별로 그 비용이 다른 값이 적용되는 반면에 대중교통의 경우는 통행목적과 상관없이 동일한 요금을 적용하게 된다. 이것은 승용차의 경우 비용적 측면에서 통행목적에 따라 선택행태가 다르게 영향을 받는 반면에 대중교통은 비교적 통행목적에 따라 비용적 영향이 비교적 큰 차이가 없다는 가정이 내재하였다고 해석할 수가 있을 것이다. 또한 다른 연구에서도 항상 어려운 문제로 제시되는 독립변수인 통행비용이 통행시간 또는 통행거리와의 상관관계가 높아 통계적으로 다중공선성 문제가 제기되는 경우가 생긴다. 이 연구에서 적용한 승용차 통행비용도 통행시간에 통행목적별 시간가치를 적용하다보니 선형적 상관관계를 갖게 되어 다중공선성

문제를 안고 있다. 하지만 Otuzar와 Willumsen(2001)에서 설명된 것과 같이 정책결정 상에 중요한 의미를 갖는 변수의 경우 다중공선성 문제를 안고라도 분석자의 판단에 따라 변수를 포함시켜 분석할 수가 있다고 본 연구에서는 판단하였다.

그 다음으로는 선택되지 않은 수단에 대한 서비스 특성에 대한 보완이다. 본 연구에서 사용한 가구통행실태 조사 자료도 다른 조사 자료와 마찬가지로 설문에 응답한 통행자가 선택한 교통수단의 서비스 속성에 대해서는 응답자가 기록을 하여 자료가 있으나 통행자별 대안집합이 무엇이며, 응답자가 선택하지 않은 다른 대안 교통수단의 서비스 속성에 대해서는 기록하지 않아 자료가 없으므로 불가피하게 기존의 자료를 가지고 추정해야 한다.

본 연구에서는 먼저, 표본자료가 충분히 많은 O-D 쌍만을 추려내어 각 O-D 쌍별로 각 교통수단별 서비스 속성의 평균값과 표준편차를 계산하였다. 이렇게 계산된 통행비용 및 통행시간과 같은 교통수단별 서비스 속성의 평균값과 표준편차를 이용하여 그 해당 i-j 쌍의 특정 통행자 n에 의해 선택된 교통수단(m)의 k번째 서비스 속성 S_{klijm}^n 의 t-value를 식(1)과 같이 계산하였다. 그리고 통행자 n이 선택하지 않은 다른 대안 교통수단 u의 서비스 속성은 식(1)에서 구한 평균값과 표준편차를 이용하여 구하였다.

$$t_{nmk}^{ij} = \frac{S_{klijm}^n - \overline{S_{klijm}}}{sd(S_{klijm})} \tag{1}$$

- 여기서, S_{klijm}^n : 통행자 n이 특정 출발점 i에서 도착점 j로 가면서 경험한 교통수단 m의 k번째 서비스 속성의 값
- $\overline{S_{klijm}}$: 출발점 i에서 도착점 j로 가는 모든 관측 통행자의 S_{klijm}^n 값의 평균 값
- $sd(S_{klijm}^n)$: 모든 관측 통행자의 S_{klijm}^n 값의 표준편차

즉, 통행자 n이 선택하지 않은 다른 대안 교통수단 u의 k번째 서비스 속성은 이미 조사 자료에서 구한

1) 이와 같은 승용차의 금전적 통행비용 추정방법은 자료의 한계에 의한 매우 개략적 방법으로 이론적 논리성이 약하므로 향후 정밀한 조사 자료에 의한 실험연구가 이루어져 좀 더 좋은 이론적 배경과 정확한 자료에 근거하여 추정할 수 있는 방법이 제시되는 것이 바람직할 것으로 고려된다.

$\overline{S_{kiju}}$ 와 식(1)에서 구한 t-value t_{nmk} 을 이용하여 식(2)와 같이 추정하였다.

$$S_{kiju}^* = \overline{S_{kiju}} + sd(S_{kiju}) \cdot t_{nmk}^{ij} \quad (2)$$

이와 같은 선택되지 않은 대안 교통수단의 서비스 속성 추정방법은 O-D 쌍 간의 속성이 출발존의 존중심(centroid)에서 도착존의 존중심으로 가는 것이 아니고 출발존 내에서 해당 통행자의 특정 출발 위치에서 도착존의 특정 도착 위치에 가는 것이기 때문에 이에 해당하는 통행비용 또는 통행시간을 다르게 추정하는 것이 바람직하다는 판단에서 비롯되었다. 즉, 공간적으로 넓은 교통존을 한 점인 존중심(centroid)으로 집합화하면서 생기는 오류를 개별 자료에 의한 모형 정산에서는 가능한 배제하고자한 노력의 일환이다.

한편, 이와 같은 추정방법 또한 정도의 차이는 있지만 역시 오류의 가능성이 있으므로 이와 같은 오류 가능성을 확률적으로 분산시키기 위해 추정된 S_{kiju}^* 의 한계 오차범위 M_k 를 논리적 판단에 의해 설정하고 그 범위 안에서 균일확률분포(uniform distribution)를 기준으로 무작위로 생성한 숫자 ϵ_k 을 생성하여(random number generation) 아래의 방법과 같이 최종적인 값으로 확정하여 모형정산에 사용하였다.

$$S_{kiju}^{*n} = S_{kiju}^* + \epsilon_k \quad (3)$$

여기서, $-M_k \leq \epsilon_k \leq M_k$

III. 모형정립

교통수단 선택모형의 정립을 위해 첫 번째 단계는 통행자들의 교통수단 선택행태를 가장 잘 설명할 수 있을 것으로 예상되는 변수의 선정과 효용함수의 형태에 대한 결정이다. 본 연구에서 변수의 선정은 기본적으로 <표 1>에서 제시된 구분과 상식적 논리성에 따라 기본 변수를 선정하였으며, 이 기본 변수와 함수형태를 기초로 하여 변수의 추가 및 제거, 함수 형태의 변형 등을 통해 기본 모형 형태보다 설명력이 높은 예측모형을 찾기 위해 통계적 검증 분석에 의한 모형 정산과정을 시행하였다. 이와 같이 더 좋은 모형을 찾기 위한 시도는 다음과 같은 기준에 의해 여러 가지 모형을 시도 분석하였다.

첫째, 모든 대안의 효용함수에서 동일한 설명변수에 대해서는 동일한 값의 계수를 적용하는 대안일반변수(alternative abstract variable 혹은 generic variable)와 동일한 설명변수라도 대안별로 다른 값의 계수를 적용하는 대안특성변수(alternative specific variable) 중 어떤 형태로 모형을 구축할 것인가를 고려하였다. 예를 들어 택시와 버스가 동일하게 500원씩 인상할 지라도 통행자의 효용 정도에 다르게 영향을 준다면 대안특성변수 형태로 효용함수를 구축해야 하고, 만일 500원 인상이 택시나 버스에 대한 선택 선호도 정도가 동일하게 영향을 준다면 대안일반변수 형태로 모형을 구축되어 질 것이다. 둘째로, 대안특성상수(mode specific constant) 설정을 몇 개의 대안에 독립적 혹은 공통적 상수 값으로 적용할 것인가를 고려하였다. 대안특성상수는 상수 값이 0으로 처리되는 기준대안과의 상대적 통행자들의 선호도를 의미한다. 즉 설명변수로는 설명되지 못한 어떤 다른 요인에 의해 대안특성상수가 있는 대안이 기준 대안보다 상대적으로 얼마나 더 선호하는가를 설명해 주는 상수 값이다. 따라서 Ben-Akiva(1985)에서 설명된 것과 같이 특정 통행자의 대안집합 수가 J개 라면 기준대안에 대한 상수 값은 0으로 설정되므로 항상 J-1개 이하의 대안특성상수를 선택모형에 사용하여야 한다. 필요에 따라서는 별도의 대안이 동일한 대안특성상수 값을 가질 수도 있으며, 또한 1개 이상의 대안이 기준대안으로 대안특성상수가 없는 0으로 처리될 수도 있다. 셋째, 사회경제변수(socioeconomic variables)를 각 대안에 어떤 형태로 포함시키는 것이 가장 적합한가를 판단하기 위한 분석을 하였다. 대안특성상수와 같이 J-1개 이하 만이 포함되어야 하는 또 다른 형태의 변수가 선택의 주체가 되는 통행자의 속성을 규명하는 사회경제변수들이다. 통행시간과 통행비용과 같이 대안 교통수단의 속성을 나타내는 변수는 대안별로 다른 값을 갖게 되나, 소득, 성별, 나이 등과 같은 통행자의 속성을 나타내는 변수는 동일한 선택 주체에 대한 대안 효용함수를 구축하는 것이기 때문에 각 대안별로 동일한 사회경제변수 값이 입력되게 된다. 따라서 이와 같은 통행자의 사회경제변수를 효용함수에 포함시킬 경우 역시 J-1개 이하만을 포함시켜야 하며 필요시에는 대안특성변수로 처리하여 J-1개 이하의 대안별로 사회경제변수에 대해 다른 계수 값을 적용할 수도 있는 것이다. 넷째, 마지막으 선택행태에 있어 독립변수의 선형적 혹은 비선형적 영향 중 적합한 함수적 영향관계를 파악 분석하였다.

〈표 2〉 효용함수식에 포함시킨 설명 변수

변수명	변수의 설명
TT	총통행시간(Total Travel Time, 단위: 분) = IVTT(차내시간) + OVTT(차외시간)
TC	총통행비용(Total Cost, 단위: 원) = 직접비용 + 간접비용
더미 변수	G 성별더미 여성일 경우, G = 0 남성일 경우, G = 1
	가구소득더미 저소득층(100만원 이하, 기준 소득계층) 소득더미1 : 중산층(101만원 이상 500만원 이하) 이면 1, 아니면 0 소득더미2 : 고소득층(501만원 이상)이면 1, 아니면 0
	A 나이더미 미성년자(20세 미만, 기준 나이 계층) 나이더미1 : 청년층(20세 이상 30세 미만)이면 1, 아니면 0 나이더미2 : 중년층(30세 이상 60세 미만)이면 1, 아니면 0 나이더미3 : 노년층(60세 이상)이면 1, 아니면 0

즉 변수들이 개인의 수단선택 효용에 선형적으로 영향을 줄지 혹은 비선형적으로 영향을 주게 될 지에 대해 상식적으로 판단해 보고 통계적으로 검증해 보는 분석이 이루어졌다. 만일 비선형적 영향을 줄 것으로 예상된다면 설명변수를 변형(transformation)시켜서 비선형적 관계를 반영시킬 수가 있다. 일반적으로 비선형관계를 반영할 때는 모형 추정의 용이성을 위해 설명변수와 종속변수와 의 관계만 비선형적으로 변화시키고, 파라메타 값을 기준으로 할 때 효용함수 기본 골격은 선형모형으로 유지하며 분석하고 있다. 즉 모형자체는 계수 값을 기준으로 선형적 모형(linear in parameters)을 유지하면서도, 종속변수와 독립변수 자체 간에는 비선형적 관계를 반영하도록 하고 있다.

본 연구에서는 자료의 한계 상 승용차를 모든 통행의 선택대안에 포함되었다고 가정하고 분석하였다. 그러나 이는 현실적이지 못한 분석 결과를 야기할 수 있으므로, 향후 기초자료 조사 시에는 대중교통 captive rider와 승용차-대중교통 choice rider를 구분하여 각 통행자 그룹별로 교통수단 대안의 집합을 다르게 구성하여 교통수단 선택모형의 정산이 이루어지는 것이 바람직하다고 생각한다.

〈표 3〉 기본모형의 효용함수식 형태

선택 대안	효용함수
승용차	$U_{승} = \beta_1 TT_{승} + \beta_2 TC_{승} + \beta_3^G G + \beta_4^I I + \beta_5^A A$
버스	$U_{버} = \beta_{비} + \beta_1 TT_{비} + \beta_2 TC_{비}$
택시	$U_{택} = \beta_{택} + \beta_1 TT_{택} + \beta_2 TC_{택} + \beta_3^G G + \beta_4^I I + \beta_5^A A$
전철	$U_{철} = \beta_{철} + \beta_1 TT_{철} + \beta_2 TC_{철}$

본 연구에서는 〈표 2〉와 같은 설명변수를 선택하고, 또한 변수들도 선형적으로 영향을 줄 것이라는 전제하에 〈표 3〉과 같이 기본모형 식을 설정한 후 모형을 추정하였다.

IV. 모형추정결과

1. 모형 1 : 기본모형

상식적이며 논리적 근거를 바탕으로 본 연구에서는 통행시간과 통행비용 그리고 통행자 속성더미를 포함하는 기본모형을 우선 설정하였다. 기본모형에서는 통행시간과 통행비용이 통행자의 효용에 모두 선형적으로 영향을 준다고 가정하고 있다. 물론 통행거리에 따라 그 영향이 다를 수는 있으나 모형의 단순화를 위해 우선 선형적 모형을 기본모형으로 설정하였다. 또한 본 연구에서는 계수 값 크기의 적절성을 검토하기 위해서 시간가치의 크기가 합리적인지도 검토하였다. 한계대체율에 의한 시간가치 분석결과, 시간 절감에 따른 지불용의액이 7,726(원/시간)이 산출되어져 최근 발표된 국토연구원(2003)에 의한 수도권 지역의 통행시간가치 7,039(원/시간)~10,218(원/시간)의 범주 안에 있어 합리적 분석 결과가 나왔다고 판단하였다. 국토연구원(2003)의 연구 결과는 대상지역이 수도권으로 시간가치가 부산·울산권 보다 높을 것으로 예상되어지고 시간가치도 2003년 기준으로 시간가치의 차이는 존재할 수가 있다고 고려된다. 이와 같이 시간가치를 계산한 것은 다른 연구의 결과와 비교하면서 상식적 판단의 값에서 동떨어진 값이 나오지는 않았는가를 검토함으로써 모형 정산과정 상에 오류를 점검하기 위한 목적이 있다. 분석결과로 볼 때 시간가치 값이 기존 연구 및 상식선에서 아주 다른 값은 나오지는 않아 모형 정산결과로 얻은 효용함수는 정책분석에 활용할 수 있다고 판단하였다.

각 독립변수에 대한 계수의 부호를 보면 총 통행시간

과 총 통행비용이 음(-) 부호가 나와 상식적인 방향과 일치하여 결과가 논리적이라고 판단된다. 또한 대안특성 상수 값은 지하철, 버스, 택시 모두 양(+) 부호가 나와 현재 효용함수 내에 포함된 설명변수에 의해 설명되지 않은 부분의 효용이 모두 승용차보다 높은 어떤 요인이 작용하고 있음을 알 수 있으며, 그 중에서도 버스가 가장 높고, 그 다음은 지하철이 높고, 그리고 택시는 승용차와 큰 차이가 없을 정도로 미미하게 높다는 것을 알 수 있다. 대안특성상수에 대해서는 상식적 선에서 부호의 방향을 판단하기가 어려우므로 그 값의 현상에 대해서만 통계적 해석만을 하여야 할 것이다. 대안특성 성별더미 변수의 계수 값을 보면 승용차의 경우나 택시의 경우 모두 양(+)의 값을 보여 남자가 여자보다 승용차 혹은 택시에 대해 더 선호적인 선택행태를 보인다고 판단되며 그 정도는 승용차가 택시보다 더 큰 것으로 분석되었다. 이 점도 어느 정도는 상식적 범위에서 설명이 되는 현상으로 효용함수 분석결과의 논리성에는 문제를 야기하지 않는다고 판단된다. 대안특성 소득더미의 경우 승용차 및 택시에 대한 중산층과 고소득층 모두가 저소득층에 비해 승용차 혹은 택시를 더 선호하는 경향을 반영하는 양(+)의 계수 부호 결과가 나왔다. 역시 중산층이나 고소득층 모두가 택시에 대한 선호도보다는 승용차에 대한 선호도가 더 높기 분석 결과가 나왔으며, 또한 고소득층이 중산층보다 택시에서나 승용차에서 모두 더 선호적인 선택행태를 하는 것으로 분석 결과가 나왔다. 이 점도 충분히 상식적 선에서 이해가 되는 논리적인 결과라고 판단된다. 마지막으로 대안특성 나이더미변수를 보면 승용차나 택시에 대해 모두 양(+)의 부호가 나와 중년층이 미성년층, 청년층 및 노년층보다 승용차와 택시에 대한 선호도가 높은 것으로 분석 결과가 나온 것이다. 택시에 대한 계수 값의 크기가 승용차의 계수 값 크기보다 적은 것으로 보아 중년층이 승용차를 다른 연령계층보다 선호하는 정도가 택시보다 큰 것으로 분석된 것이다. 이것도 상식적으로 판단하는 방향과 일치하여 논리적인 분석 결과가 나왔다고 판단된다.

택시 대안특성변수 성별더미의 추정 계수 값에 대한 t 값이 1.8인 것을 제외하고 모든 변수의 t 값이 2의 값보다 커서 매우 높은 통계적 유의성을 보이고 있어 모든 계수 값이 0과 다를 확률은 매우 높다고 해석할 수 있다. 이것은 모든 설명변수가 모형에 포함되어 통행자의 교통수단 선택행태를 설명하는데 통계적으로 중요한 변수라는 것을 의미하기도 하는 것이다. 성별더미의 경우도

1.8로 95% 신뢰도에서 일반적 검증 기준이 t 값 1.96 보다는 적지만 그렇다고 통계적 의미성이 없다고 할 정도로 적은 것은 아니므로 충분히 모형 안에 포함될 수 있다고 생각된다.

기본모형에 의한 모형 전반적인 수단선택 행태의 설명력을 판단하기 위해 ρ^2 값을 사용할 수 있다. 상수 포함 모든 계수 값이 0일 때의 우도함수(Log-likelihood function) 값을 기준으로 계산한 $\rho^2(0)$ 값이 0.5 정도로 윤대식(2001)에서 서술한 규모 이상으로 충분히 높은 값이 나와 만족할 만한 결과라고 판단된다. 다만 대안특성상수의 설명력을 뺀 즉 상수 값만 0이 아닌 추정된 값이 적용되고 다른 모든 계수 값이 0일 때의 우도함수 값을 기준으로 계산한 $\rho^2(C)$ 이 0.26 정도로 $\rho^2(0)$ 값 0.5보다는 작아 모형 설명력 중에 많은 부분이 상수 값에 의해 설명되고 있음을 알 수 있다.

<표 4> 부산·울산권 기본모형의 계수 추정

설명변수 명	추정 계수 값	Std. error	t 값
총 통행시간 (분)	-1.62E-02	1.51E-03	-10.7
총 통행비용 (원)	-1.258E-04	1.14E-05	-11.1
지하철 대안특성 상수	1.163	0.114	10.2
버스 대안특성 상수	3.349	0.095	35.4
택시 대안특성 상수	0.500	0.180	2.8
승용차 대안특성변수 성별더미 (남성)	1.829	0.061	30.2
승용차 대안특성변수 소득더미1 (중산층)	0.958	0.072	13.3
승용차 대안특성변수 소득더미2 (고소득층)	1.308	0.164	8.0
승용차 대안특성변수 나이더미2 (중년층)	3.091	0.070	44.5
택시 대안특성변수 성별더미 (남성)	0.214	0.116	1.8
택시 대안특성변수 소득더미1 (중산층)	0.294	0.141	2.1
택시 대안특성변수 소득더미2 (고소득층)	0.907	0.291	3.1
택시 대안특성변수 나이더미2 (중년층)	1.597	0.127	12.6

<통계치 요약>
 관측 표본수 = 9,236
 $L(0) = -12,804$
 $L(c) = -8,398$
 $L(\beta) = -6,190$
 $-2[L(0) - L(\beta)] = 13,227$
 $\rho^2(0) = 0.5165$
 $\rho^2(C) = 0.2629$
 $\bar{\rho}^2(0) = 0.5155$

2. 모형 2 : 개선모형 -1

좀 더 설명력이 높은 교통수단선택 모형을 찾기 위해 기본모형을 약간 변형한 개선모형-1을 시도하여 보았다. 기본모형에서 바뀐 부분은 통행비용 변수를 계수 값이 승용차와 대중교통(지하철, 버스, 택시)에 다르게 적용하는 대안특성변수로 바꾸어 분석한 것이다.

개선모형-1의 분석 결과를 보면 각 설명변수에 대한 계수의 부호와 크기는 앞에서 기본모형에서와 마찬가지로 논리적으로 합리적인 결과가 산출되었다. 다만 통행비용이 승용차와 대중교통에 대한 만족도의 정도를 다르게 설명하도록 효용함수를 설정하면서 비용 외의 다른 설명변수 계수 값이 약간씩 변화가 생겼지만 심각한

차이는 없어 보인다. 이와 같은 변화는 설명변수 간의 상관관계수가 0이 아님으로써 새로운 변수에 의해 설명되는 부분이 상관관계 부분에 대한 상호 상쇄되는 현상에 의한 것으로 문제시될 필요는 없다고 판단된다. 통행비용의 계수 값 크기를 보면 대중교통의 계수값이 승용차의 계수값 보다 절대 값이 더 크므로 대중교통의 요금이 대중교통 선호도에 더 민감하게 영향을 준다는 분석 결과가 나온 것이다. 이러한 현상은 일반적으로 인식하고 있는 승용차 이용자가 비용에 덜 민감하다는 상식적 판단과 일치한다.

계수 값 크기에 있어서는 기본모형의 택시 대안특성 상수 0.4995가 개선모형-1에서는 3.306으로 가장 큰 변화를 보이는데 이것은 통행비용의 효용함수에 미치는 영향을 승용차와 대중교통으로 분리하면서 택시가 대중교통의 범주에 포함되어 지하철, 버스의 영향이 간접적으로 상수에서 포함되었기 때문이다. t 통계치를 보면 택시 성별더미의 계수에 대한 t 값이 기본모형의 1.8 보다 높은 3.1로 높아져 그 통계적 유의성이 훨씬 좋아졌으나, 택시에 대한 중산층의 소득더미는 2.1에서 1.5로 상대적으로 통계적 유의성이 많이 떨어져 90% 이상의 확신을 갖고 그 계수 값이 0이 아니라고 통계적으로 단언하기가 어렵다. 그 외의 다른 모든 설명변수의 통계적 유의성은 모두 매우 높게 분석되어 만족할 만한 분석 결과가 나왔다. 따라서 택시의 소득더미 변수를 제외하고는 각 독립변수별 통계적 의미성 측면에서 적합하여 정책분석에 활용 가능한 모형이라고 판단된다.

모형의 전체 설명력 측면에서 기본모형 $\rho^2(0)$ 값이 0.5165에서 개선모형-1인 경우 0.5445로 향상되었다. 설명변수 수의 증가를 고려하더라도 기본모형의 $\rho^2(0)$ 값은 0.5155이고 개선모형-1은 0.5434로 향상되었다.

<표 5> 부산·울산권의 개선모형-1의 계수 추정

설명변수 명	추정 계수 값	Std. Error	t 값
승용차 대안특성변수 통행비용 (원)	-7.576E-05	1.26E-05	-6.0
대중교통대안특성변수 통행비용 (원)	-8.599E-04	3.72E-05	-23.1
총 통행시간 (분)	-0.011	0.002	-7.4
지하철 대안특성 상수	1.827	0.118	15.5
버스 대안특성 상수	3.797	0.097	39.2
택시 대안특성 상수	3.306	0.215	15.4
승용차 대안특성변수 성별더미 (남성)	1.768	0.060	29.7
승용차 대안특성변수 소득더미1 (중산층)	0.955	0.072	13.3
승용차 대안특성변수 소득더미2 (고소득층)	1.106	0.164	6.8
승용차 대안특성변수 나이더미2 (중년층)	2.860	0.069	41.8
택시 대안특성변수 성별더미 (남성)	0.382	0.123	3.1
택시 대안특성변수 소득더미1 (중산층)	0.230	0.151	1.5
택시 대안특성변수 소득더미2 (고소득층)	0.865	0.315	2.7
택시 대안특성변수 나이더미2 (중년층)	1.706	0.134	12.7
<통계치 요약> 관측 표본수 = 9236 $L(0) = -12803.8147$ $L(c) = -8397.7586$ $L(\beta) = -5831.6413$ $-2[L(0) - L(\beta)] = 13944.3468$ $\rho^2(0) = 0.5445$ $\rho^2(C) = 0.3056$ $\rho^2(0) = 0.5434$			

3. 모형 3 : 개선모형 -2

개선모형-1보다 통계치가 좋으며, 모형 설명력을 더 높일 수 있는 모형을 찾기 위해 개선모형-2와 같이 효용함수 형태를 변형하였다. 개선모형-1에서 택시에 대한 중산층의 소득더미 계수가 0이 아니라고 하기에는 통계적으로 유의성이 떨어지므로 개선모형-2에서는 제외시켰으며, 또한 교통수단 선택행태에 있어 통행자의 통행목적과 통행자 직업이 영향을 줄 것으로 예상되

므로 대안특성변수로써 출퇴근 통행목적 더미를 승용차와 지하철에 각각 포함시켰다. 또한 3차 산업 종사자들이 다른 직업 종사자에 비해 승용차 선호도가 높을 것이라는 가정에 따라 3차 산업 종사자에 대한 승용차 대안특성 더미변수를 개선모형-2에 포함시켜 분석하였다.

개선모형-2의 분석 결과를 검토하면 계수의 부호, 계수 값의 상대적 크기 등이 앞에서 설명한 기본모형과 개선모형-1과 같이 상식적으로 수긍할 수 있는 논리적인 결과가 제시되었다. 각 설명변수의 t 값은 모두 절댓값이 2 이상으로 상당히 높은 통계적 유의성을 보여

기본모형 및 개선모형-1보다 상대적으로 좋은 결과가 나왔다. 새로 추가된 통행목적 더미변수와 직업더미변수에 대한 계수 값이 의미하는 것을 해석하면, 출퇴근 통행목적의 경우 승용차(계수값=0.8568)와 전철(계수값=0.4798)을 선호하는 정도가 다른 통행목적에 비해 높으며, 또한 출퇴근 시에 승용차를 전철보다는 더 선호하는 것으로 분석 결과가 산출되었다. 또한 승용차 대안특성변수로 3차 산업 직업(회사원, 공무원, 서비스직)의 직업더미에 대한 계수 값이 양의 값을 갖는 0.3599로 분석되어 3차 산업 직업을 갖는 통행자들이 다른 직업을 가진 통행자들보다 승용차를 더 선호하는 경향이 있음을 보여주고 있다. 이것도 개선모형-2를 설정하면서 양의 부호로 나올 것을 기대하고 변수를 추가 한 것이었으므로 상식적으로 판단한 통행 형태와 같은 방향이라고 판단된다. 따라서 각 설명변수의 부호, 계수의 값 및 변수의 논리적 설명력 등이 적합하게 분석 결과가 나왔다.

모형 전반적인 설명력 측면에서도 개선모형-2가 기본모형이나 개선모형-1보다 향상된 것을 볼 수 있다. 즉 $\rho^2(0)$ 값이 0.5528로 기본모형의 0.5165, 개선모형-1의 0.5445보다 높게 분석 결과가 나왔으며, $\overline{\rho}^2(0)$ 의 경우도 0.5516이 나와 기본모형의 0.5155, 개선모형-2의 0.5434보다 높게 나와 추가적인 변수로 인해 잃어버린 자유도(degree of freedom)보다는 더 설명력이 향상되었음을 알 수 있다. 또한 대안특성상수의 설명력을 제외시키기 위해 대안특성상수의 값만 추정된 값을 쓰고 나머지 설명변수에 대한 계수 값은 0으로 놓고 계산한 우도함수값을 기준으로 한 $\rho^2(C)$ 값도 0.3182으로 기본모형의 0.2629, 개선모형-1의 0.3056보다 크게 나와 설명변수에 의한 설명력도 향상되었음을 알 수 있다.

〈표 6〉 부산·울산권의 개선모형-2의 계수 추정

설명변수 명	추정 계수값	Std. Error	t 값
승용차 대안특성변수 통행비용 (원)	-1.828E-04	1.51E-05	-12.1
대중교통 대안특성변수 통행비용 (원)	-9.409E-04	3.85E-05	-24.4
총 통행시간 (분)	-7.328E-03	1.54E-03	-4.7
지하철 대안특성 상수	1.437	0.127	11.3
버스 대안특성 상수	3.621	0.096	37.8
택시 대안특성 상수	3.463	0.189	18.4
승용차 대안특성변수 성별더미 (남성)	1.616	0.060	26.9
승용차 대안특성변수 소득더미1 (중산층)	0.895	0.071	12.7
승용차 대안특성변수 소득더미2 (고소득층)	1.027	0.163	6.3
승용차 대안특성변수 나이더미2 (중년층)	2.470	0.072	34.4
택시 대안특성변수 성별더미 (남성)	0.484	0.125	3.9
택시 대안특성변수 소득더미2 (고소득층)	0.682	0.297	2.3
택시 대안특성변수 나이더미2 (중년층)	1.833	0.137	13.4
승용차 대안특성변수 통행목적더미(출퇴근통행)	0.857	0.079	10.8
지하철 대안특성변수 통행목적더미(출퇴근통행)	0.480	0.141	3.4
승용차 대안특성변수 직업더미 (3차 산업)	0.360	0.066	5.4

〈통계치 요약〉

관측 표본수 = 9236
 $L(0) = -12803.8147$
 $L(c) = -8397.7586$
 $L(\beta) = -5725.4643$
 $2[L(0) - L(\beta)] = 14156.7008$
 $\rho^2(0) = 0.5528$
 $\rho^2(C) = 0.3182$
 $\overline{\rho}^2(0) = 0.5516$

4. 모형 4 : 네스티드 로짓모형

1) 네스티드 로짓모형

다항로짓모형(Multinomial Logit Model)의 가장 단점으로 지적되고 있는 “비관련 대안으로부터의 독립성(IIA, independence from irrelevant alternatives)”이라는 로짓모형 속성에 대해서는 이미 잘 알려진 내용이다. 교통수단선택 모형에 있어 가장 많이 논의되는 것이 승용차와 대중교통은 독립적으로 볼 수가 있더라

도 대중교통 내에서의 수단 간에는 설명변수로 설명이 되지 않는 오류부분의 분포가 상호 연관성이 존재할 수 있다는 생각에서 IIA 속성의 문제점이 거론되고 있다. 따라서 본 연구에서는 통행실태조사 자료가 다항로짓모형의 IIA 속성에 적합한지 혹은 네스티드 로짓모형으로 분석해 보는 것이 바람직한 것인가를 먼저 검증하고, 만약 네스티드 로짓모형 구조가 더 타당하다면, 네스티드 모형을 추정하는 것이 좋을 것이다.

네스티드 로짓모형으로 다양한 위계구조를 시도할 수가 있겠지만 본 연구에서는 상위구조(Upper level)는 승용차와 대중교통의 선택구조로 하고, 하위구조(Lower level)는 대중교통으로써 전철, 버스, 택시의 선택구조로 구성하여 분석하였다. 먼저, Chi-square 검정을 통한 네스티드 모형 구축의 당위성을 검증하였다. 개별통행자료에서 모든 교통수단 대안이 모두 포함된 자료를 U(Universal choice set)라고 하고, 확률변수 ϵ 간에서 상관관계가 있을 것으로 추정되는 대안의 부분집합의 교통수단을 선택한 통행자들의 개별자료만 추려서 모은 개별자료를 S(Subset of alternative)라고 정의하였다. 만일 통행실태 조사자료가 다항로짓모형의 IIA 속성을 만족시킨다면 U 자료를 이용하여 구축한 모형과 S 자료를 이용하여 구축한 모형의 각 독립변수의 계수 값이 유사하게 제시되어야 한다. 그러나 이 값이 매우 차이가 많이 날 경우에는 현재 사용하고 있는 통행자료가 적용된 다항로짓모형의 IIA 속성을 만족시키지 못하고 있다고 판단하게 된다.

Small and Hsiao(1982)가 제안한 아래와 같은 Chi-square test를 통해 이를 검증할 수 있다. 귀무가설은 U 자료에서 구한 효용함수 내의 계수 값과 S 자료에서 구한 계수 값이 동일하다는 가정을 놓고 통계적 검증을 하는 것이다.

$$\text{귀무가설 } (H_0) : \overline{\beta_U} = \overline{\beta_S} \quad (4)$$

chi-square 값 :

$$\chi^2 = - \frac{2[LL_S(\beta_U) - LL_S(\beta_S)]}{1 - N_S/(aN_U)} \quad (5)$$

여기서, 자유도(d.f) : 자료 S를 이용한 모형 내의 계수 수

N_U : U 의 표본 수

N_S : S 의 표본 수

α : scalar ($\alpha > 1$ 인 숫자이지만 일반적으로 $\alpha = 1$ 을 적용)

$LL_S(\beta_U)$: U 자료를 이용하여 구한 계수 값을 이용하여 S자료에서 구한 우도함수 값

$LL(\beta_S)$: S 자료를 이용한 우도함수 값

부산·울산권의 자료에서 승용차를 선택한 자료를 모두 제거하고 대중교통을 선택한 자료만 구축하여 S 자료라고 설정하였으며, 모든 자료가 포함된 자료를 U 자료로 한 후 위의 식을 계산하였다. 위와 같은 방법으로 계산하여 구한 χ^2 값이 548.5이며, 반면에 자유도가 8이고 유의수준(significant level) α 가 0.05인 임계치(critical value) χ^2 값은 15.5에 불과하다. 이것은 U 자료에서 구한 계수 값과 S 자료에서 구한 계수 값이 95%보다도 훨씬 높은 확률의 확신을 갖고 귀무가설을 기각할 수 있다. 따라서 부산·울산권의 자료는 승용차와 대중교통이 상위선택위계 그리고 대중교통 가지 밑으로는 전철, 버스, 택시의 교통수단이 하위선택위계 구조를 가진 네스티드 로짓모형 구조를 가질 가능성이 높다고 판단할 수가 있다.

2) 추정결과

부산·울산권의 개별통행실태 자료를 이용하여 네스티드 로짓모형에 대해 모수 추정을 하고 앞에서 분석하여 가장 우수한 모형으로 선정된 다항로짓모형인 개선모형-2와 비교 검토하여 보았다. 앞에서 구축한 개선모형-2의 효용함수 형태를 그대로 적용하고, 네스티드 구조는 앞에서 서술한 바와 같이 상위 선택구조에는 승용차와 대중교통(composit alternatives의 개념)으로 설정하고, 대중교통 가지 밑으로 전철, 버스 및 택시로 하위 선택구조를 설정하여 분석하였다.

네스티드 로짓모형의 분석 결과와 다항로짓모형이 개선모형-2와 비교해 보면 계수의 부호, 크기 및 설명의 논리성, t 값 등 두 모형 모두 합리적인 결과가 나왔으며 계수 값의 차이도 크게 나지 않는다.

계수 값의 크기를 비교할 때는 대중교통의 집합에서 계산된 효용 최대값의 기대값인 Inclusive value(혹은 Log-sum value) 앞에 붙는 계수인 θ (THETA) 값이 0.6918이라는 값의 크기가 중요하다. 즉 네스티드 로짓모형의 상위구조와 하위구조의 다른 분산 값을 표준

화(normalization)하는 과정에서 동일한 분산 값을 갖도록 조정된 값이 θ 이므로 대중교통의 계수 값이 개선모형-2보다 약간씩 크게 추정되는 것은 이해될 수 있는 부분이다. 이론적으로 θ 값은 $0 \leq \theta \leq 1$ 사이의 값을 갖도록 되어 있다. 즉 θ 값이 0에 가까우면 하위구조의 그룹 내의 대안간 ϵ 의 상관계수가 매우 높아 네스티드 로짓모형 구조라야 오류가 없으며, 만일 θ 값이 1에 가까우면 대안간 ϵ 이 매우 독립적이므로 일반적인 다항 로짓모형을 적용하여도 큰 문제가 없다고 해석할

수가 있다. 부산·울산권 분석의 경우에는 θ 값이 0.6918로 대중교통수단간에 독립적이라고 보기가 어려울 것으로 보인다. 통계적으로 분석해 보면 귀무가설 $H_0 : \theta = 1$ 로 놓고 t-test를 하면, t 값은 $t = (0.6918 - 1) / 0.0604 = -5.103$ 으로 매우 높은 확률을 가지고 θ 는 1이 아니라고 주장할 수가 있다. 따라서 부산·울산권의 교통수단선택 모형은 네스티드 로짓모형의 구조가 더 바람직한 것으로 판단된다.

한편, 모형 전체적 설명력을 나타내는 ρ^2 값을 비교한 결과, 개선모형-2의 $\rho^2(0)$ 의 값은 0.5528이고, $\rho^2(0)$ 의 값은 0.5516인데 네스티드 로짓모형의 경우는 각각 0.5537과 0.5524으로 네스티드 로짓모형이 개선모형-2보다 향상은 되었으나, 모형의 복잡성에 비해 향상시킨 값은 미미한 수준이라고 판단된다.

<표 7> 부산·울산권의 네스티드 로짓모형의 계수 추정

설명변수 명	추정 계수값	Std. Error	t 값
승용차 대안특성변수 통행비용 (원)	-1.22E-04	2.03E-05	-6.0
대중교통 대안특성변수 통행비용 (원)	-1.04E-03	4.66E-05	-22.2
총 통행시간 (분)	-8.30E-03	1.64E-03	-5.1
지하철 대안특성 상수	2.858	0.427	6.7
버스 대안특성 상수	5.059	0.423	12.0
택시 대안특성 상수	5.071	0.482	10.5
승용차 대안특성변수 성별더미 (남성)	1.592	0.059	27.0
승용차 대안특성변수 소득더미1 (중산층)	0.891	0.070	12.7
승용차 대안특성변수 소득더미2 (고소득층)	1.053	0.159	6.6
승용차 대안특성변수 나이더미2 (중년층)	2.436	0.071	34.3
택시 대안특성변수 성별더미 (남성)	0.552	0.135	4.1
택시 대안특성변수 소득더미2 (고소득층)	0.808	0.326	2.5
택시 대안특성변수 나이더미2 (중년층)	1.877	0.144	13.0
승용차 대안특성변수 통행목적더미(출퇴근통행)	0.686	0.088	7.8
지하철 대안특성변수 통행목적더미(출퇴근통행)	0.508	0.143	3.6
승용차 대안특성변수 직업더미 (3차 산업)	0.356	0.066	5.4
θ (THETA, Inclusive Value - 대중교통)	0.692	0.060	11.5
<통계치 요약> Number of observations = 9236 L(0) = -12803.8147 L(c) = -8397.7586 L(β) = -5714.5779 2[L(0) - L(β)] = 14178.4736 $\rho^2(0)$ = 0.5537 $\rho^2(C)$ = 0.3195 $\rho^2(0)$ = 0.5524			

V. 장래 교통수단 분담율의 예측

지금까지는 관측된 통행자들의 교통수단선택 행태 자료로써 수단선택모형 구축결과를 제시하였다. 그러나 모형의 구축이 현재에 관측된 자료로부터 통행자의 교통수단선택 행태를 얼마나 잘 설명하는가도 중요하지만 교통 시스템 또는 사회경제의 변화에 따른 장래의 통행패턴이 어떻게 될 것인가를 예측하는데 모형 구축의 주요 목적이 있다. 따라서 관측된 자료로부터 가장 현실적 현상을 잘 설명하는 교통수단 선택모형이 구축되었으면, 구축된 모형으로 장래 변화를 어떻게 반영하여 미래 통행패턴을 예측할 것인가에 대한 예측기법이 명확하게 정립될 필요가 있는 것이다.

일반적으로 관측자료를 이용하여 모형을 정산할 경우에는 개별 통행자별로 통행 및 사회경제 속성에 대한 자료가 모두 존재하여 개별자료(disaggregate data)를 이용한 모형정산이 가능하지만, 장래 예측에 있어서는 각 통행자 개별적인 통행자 속성 및 교통수단별 서비스 속성을 예측하는 것은 불가능하다. 따라서 예측 시에는 개별적 자료로 독립변수가 입력될 수가 없으며, 다만 평균적 개념의 집합적 분석이 불가피한 것이다. 즉 정산 시에는 개별자료를 입력자료로 하여 정산하여 구축된 모형에다가 예측 시에는 집합적 자료의 평균값을 넣고 예측하여야 한다. 이때 모형 정산시의 자료형태와 예측시의 자료형태가 다르게 되기 때문에 예측분석 결과인 교통수단 분담율에 왜곡이 발생할 수가 있

다. 즉 선형회귀분석과 같이 선형모형의 경우에는 문제가 발생하지 않으나, Otuzar and Willumsen(2001)에서 설명된 것과 같이 로짓모형과 같이 비선형모형의 경우에는 Jensen's inequality 문제가 발생되어 예측시에 오류가 발생할 수가 있다. 이와 같은 문제점을 해결하기 위한 방법으로 여러 가지 예측기법들이 학자들에 의해 제안되어 왔다(Manhiem, 1979). 본 연구에서는 현실적으로 가장 보편적으로 활용할 수 있는 분류(classification) 방법에 의한 예측 기법을 활용한다. 분류(classification) 예측기법은 소집단으로 특성에 따라 구분하고(market segmentation), 각 특정 소집단의 속성 평균값을 개별수요모형에 적용하여 소집단의 대표적 통행자의 수요를 구하고, 대표적 수요를 해당 소집단의 통행자수로 곱함으로써 각 소집단에서의 교통수요를 예측하는 방법이다.

분석의 용이성을 고려하여 네스티드 모형을 적용하지 않고 좀 더 모형형태가 단순한 기본모형을 이용하여 교통수단 분담율을 예측하였다. 정산된 모형의 정확성을 검증하는 차원에서 표본자료를 활용하여 통행자의 속성을 기준으로 성별(남, 여), 소득(저소득층, 중산층, 고소득층), 나이(중년층, 그 외의 연령층)로 2×3×2=12개의 계층(c)으로 표본자료를 분류하여 계층별 총 통행자수(OD_{ij}^c)를 파악하였다. 모든 통행자들의 개별적 위치와 대중교통 접근성을 고려한 통행시간과 통행비용을 예측 시에는 입력하기가 매우 어렵다는 점을 감안하여 특정 O-D간의 각 교통수단별 한 값만 제공하는 평균값의 개념으로 입력한다고 가정하였다. 따라서 각 O-D별 표본의 평균 통행시간과 통행비용을 예측모형에 적용하였다. 또한 효용함수의 성별더미, 소득더미 및 나이더미와 같은 더미변수는 각 계층별 소집단에 해당되는 적합한 0과 1의 각 더미변수 값을 입력하여 적용하였다. 즉, 교통존 i에서 j로 가는 통행자 계층 c의 총 통행수요(OD_{ij}^c) 중에 교통수단 m을 선택한 수요는 식(6)과 같이 계산될 수 있다.

$$OD_{m/ij}^c = OD_{ij}^c * P_{m/ij}^c \quad (6)$$

이 수요를 모든 계층에 대해 더하면 교통존 i에서 j사이 교통수단 m을 이용하는 총 수요(OD_{m/ij})는 식(7)과 같이 계산된다.

〈표 8〉 부산·울산권 표본 수단분담율과 모형예측 수단분담율 비교

권역별 수단분담율 \ 교통수단	지하철	버스	승용차	택시
표본의 수단분담율	2.73%	35.75%	57.65%	3.87%
모형 계산에 의한 수단분담율	2.07%	36.30%	58.02%	3.61%

$$OD_{m/ij} = \sum_c OD_{m/ij}^c \quad (7)$$

따라서 교통존 i에서 j의 교통수단 m 분담율(P_{m/ij})과 도시전체에서 교통수단 m의 분담율(P_m)은 다음과 같이 계산된다.

$$P_{m/ij} = \frac{OD_{m/ij}}{\sum_m OD_{m/ij}}, \quad P_m = \frac{\sum_i \sum_j OD_{m/ij}}{\sum_i \sum_j \sum_m OD_{m/ij}} \quad (8)$$

앞에서 예측을 위한 입력자료에 의해 각 P_{m/ij}^c를 계산하고 위에서 설명한 계산 과정을 거쳐 부산·울산권의 교통수단 분담율을 계산하였다. 그리고 정산된 기본모형이 집합적 교통수단 분담율을 얼마나 정확하게 예측하였는가를 판단하기 위해 모형에 의한 예측 분담율과 실제 표본의 분담율을 비교하였으며, 그 분석 결과는 〈표 8〉에 제시되어 있다. 〈표 8〉에서 알 수 있듯이 정산된 교통수단선택모형이 도시전체의 집합적 교통수단 분담율을 대체로 정확하게 예측하고 있음을 알 수 있다.

VI. 결론

본 연구에서는 부산·울산 광역권의 통행행태 개별자료를 이용한 로짓모형 형태의 교통수단선택 모형의 정산 및 구축에 대한 상세하고 투명하게 제시를 통해 부산광역시 교통정책에 실질적인 적용이 가능한 예측모형을 제시하였다. 즉 수단선택모형 구축을 위한 조사방법, 변수설정, 모형정산, 예측기법, 입력자료 예측 등의 일련의 과정을 투명하고 명확하게 설명함으로써 제 3의 분석자도 본 연구와 같이 동일한 분석이 가능하고 본 논문의 내용을 기초로 더욱 발전적 모형이 개발될 수도 있을 것이라 생각한다. 그럼으로써 교통전문가 간에 활발한 논의의 동기가 되어 지속적인 실무적 정책분석의 발전이 있을 것으로 기대된다. 본 연구에서는 선

택되지 않은 교통수단에 대한 자료가 일반적으로 없다는 점을 보완하기 위해 비선택대안의 속성 추정방법의 제안을 하였으며, 교통수단선택 모형 정산 시의 복합 교통수단 이용에 대한 처리방법, 통계적 검증을 통한 더욱 우수한 예측모형을 찾아가는 방법, 모수 추정 결과에 대한 통계적 해석 방법, 예측 시의 집합화 오차를 줄이는 방법 등에 대해 구체적이며 명료하게 설명함으로써 교통전문가들 사이의 향후 교통수단선택 모형 정산 및 예측에 있어 주요 주제에 대해 발전적으로 토론할 수 있는 계기를 마련하고자 하였다.

향후 연구내용으로는 설명변수의 보다 정확한 추정과 종속변수의 정확한 설정을 들 수 있다. 예를 들어 승용차의 통행비용인 경우 향후 정밀한 조사를 통한 추정방법의 개발이 필요하며 본 연구에서는 자료의 한계 때문에 승용차를 모든 통행의 선택대안에 포함시켜 분석하였으나, 향후에는 기초 자료 조사단계에서부터 좀 더 정교하게 구축되어져서 대중교통 고정승객(transit captive rider)과 선택적 통행자(choice rider)를 구분하여 각 통행자 그룹별로 교통수단 대안의 집합이 다르게 구성하여 교통수단 선택모형의 정산이 이루어져야 할 것이다.

참고문헌

1. 국토연구원(2003), "유료도로의 통행시간가치 산정에 관한 연구 - 수도권 지역을 중심으로-", 연구보고서.
2. 교통개발연구원(2000), "1999년 전국교통 DB구축사업 - 가구통행실태조사 및 기초분석-", 연구보고서.
3. 교통개발연구원(2002), "2001년 전국교통 DB구축사업 - 개인통행실태조사 : 5개광역시 광역권 -", 연구보고서.
4. 김대웅·배영석·이명미(1999), "취업자의 1일 통행수

단선택 분석 및 모형의 시간이전성 검토", 대한교통학회지, 제17권 제5호, 대한교통학회, pp.19~32.

5. 서울시정개발연구원(2001), "자동차 소유, 운행의 사적·사회적 비용 비교 연구", p.59.
6. 윤대식(1997), "통근통행자의 통행패턴 선택행태의 분석", 대한교통학회지, 제15권 제4호, 대한교통학회, pp.35~51.
7. 윤대식(2001), "교통수요분석론", 박영사.
8. 조중래·김채만(1998), "출근통행 교통수단 선택행태의 지역간 비교연구 : -서울과 일산신도시를 중심으로-", 대한교통학회지, 제16권 제4호, 대한교통학회, pp.75~86.
9. 한국개발연구원(1999), "도로 및 철도부문의 예비타당성조사 표준지침 연구".
10. Ben-Akiva, Moshe and Lerman, Steven (1985), "Discrete Choice Analysis : Theory and Application to Travel Demand", The MIT Press.
11. FHWA(1977), "Computer Programs for Urban Transportation Planning, PLANPAC/BACKPAC General Information Manual", US Department of Transportation, Federal Highway Administration.
12. Manheim(1979), "Fundamentals of Transportation Systems Analysis, Volume 1 : Basic Concepts", The MIT Press.
13. Ortuzar, Juan and Willumsen, Luis(2001), "Modelling Transport", John Wiley & Sons, Ltd.
14. Small, K. and C. Hsiao(1982), "Multinomial Logit Specification Tests", Working paper, of Economics, Princeton University, Princeton, N.J.

✉ 주 작성자 : 김익기
 ✉ 논문투고일 : 2004. 11. 18
 논문심사일 : 2004. 12. 14 (1차)
 2005. 3. 23 (2차)
 2005. 5. 24 (3차)
 2005. 6. 7 (4차)
 심사판정일 : 2005. 6. 7
 ✉ 반론접수기한 : 2005. 10. 31