

O/D 및 P/A 통행량 전수화와 신뢰성 확보 방안



김익기

I. 머리말

우리나라의 교통체계효율화법에 따라 매 5년마다 정기교통조사를 실시하여 함으로 여객실태조사를 통한 기종점통행량(O/D) 자료를 구축하도록 되어있다. 또한 지방자치제도가 우리나라에서 점차 정착되어 가면서 각 광역자치단체 및 기초자치단체에서도 과학적이고 합리적인 교통정책수립을 위해 교통기초자료의 필요성을 인식하기 시작하고 있으므로 자치단체별 상황에 적합한 교통조사가 점차 확대되리라 예상된다. 이와 같은 교통조사 자료 가운데 가구통행실태 표본조사를 통해 표본 O/D (Origin-Destination) 및 P/A (Production-Attraction) 통행량 패턴을 파악하고, 이와 같은 표본 자료를 이용하여 시스템 전체의 패턴을 추정 파악하는 O/D 및 P/A의 전수화 분석 과정이 필수적으로 필요하게 된다. O/D와 P/A 개념은 김익기(1997)에 상세히 설명되어 있고, O/D와 P/D 통행 자료의 분석 상 구체적 활용은 김익기(2006)에 설명되어 있으므로 이 글에서는 개념과 활용에 대한 설명은 생략하였다. 가구통행실태조사와 O/D 및 P/A 전수화 분석은 여러 가지 교통조사와 분석 가운데 가장 경제적, 시간적 비용이 많이 드는 조사사업 중에 하나이며, 또한 자료의 정확성과 신뢰성을 확보하는데 많은 어려

움이 있는 조사이며 분석 과정인 것이다. 그리고 과학적 교통정책 수립을 위해 가장 기초적으로 필요한 입력 자료가 바로 전수화 O/D 및 P/A 통행 자료인 것이다. 따라서 이와 같은 전수화 O/D 및 P/A 통행량 자료를 구축하기 위한 가구통행실태 조사사업에서부터 전수화 분석과정을 가능한 합리적으로 수행하여 신뢰성이 높은 자료를 구축하는 것은 과학적이고 합리적인 교통정책분석을 위한 중요한 기초 과제인 것이다. 따라서 이 글에서는 O/D 및 P/A 전수화의 이론적 개념에 대해 논의하고, 교통 조사단계에서 전수화 분석단계에 이르기까지 최종 자료의 신뢰성을 높이기 위한 방법과 고려해야 할 사항에 대해 논의하고자 한다.

II. 전수화(Expansion, Grossing-up)의 기본 개념

전수화의 개념은 실질적으로 매우 단순한 개념이다. 즉 '표본조사 자료를 기반으로 1개의 표본자료가 모집단(population)내에서 대표하는 자료의 숫자(전수화 계수, expansion factor)만큼 표본자료를 확대하여 모집단 전체의 총량적 교통패턴을 가능한 유사한 값으로 추정하는 과정이다'라고 정의할 수 있을 것이다. 즉 가장 단순한 경우를 예로 들면 표본조사가 완벽히 무작위 표본추출(simple random sampling) 되었다면 표본을 역수를 표본자료에 곱함으로써 단순히 전수화 값을 추정할 수 있다. 이 경우는 조사설계(survey design)와 표본설계(sample design)의 부적절성에 의한 표본오류(sampling bias)가 존재하지 않는다는 전제가 내재하고 있는 것이다. 이와 같이 가장 단순한 경우의 전수화 과정을 수식으로 표현하면 아래와 같다.

$$f = \frac{\sum_n^N X_n^p}{\sum_k^K X_k^s} \quad (1)$$

$$Y^p = f \left\{ \sum_k^K Y_k^s \right\} \quad (2)$$

여기서 f : 전수화 계수 (expansion factor)

X_n^p : 모집단 n번째 자료의 전수화 비교변수 X의 값 ($n=1, 2, \dots, N$)

X_k^s : 표본 k번째 자료의 전수화 비교변수 X의 값 ($k=1, 2, \dots, K$)

Y_k^s : 변수 X와 상관관계가 있는 표본 k번째 자료의 전수화 대상변수 Y의 값

Y^p : 전수화 대상 변수 Y의 전수화 추정 값

위의 식(1)에서 볼 수 있듯이 전수화 계수를 계산하기 위해서는 모집단 전수가 조사된 자료의 변수 X와 표본조사 자료의 변수 X가 전수 자료와 표본 자료에 함께 있어야 만이 전수화 계수(expansion factor)의 계산이 가능하다. 이와 같은 변수 X를 이 글에서는 '전수화 비교변수'라고 하였다. 표본 O/D 및 P/A 통행의 전수화 과정에서 교통존의 가족수별 가구 수, 자동차 보유대수별 가구 수, 주택형태별 가구 수 같은 자료가 주택센서스조사와 가구통행실태 표본조사에서 함께 포함되어 있어 전수화 비교변수의 예가 될 수 있다. 또한 식(2)에서 추정하고자 하는 Y 값은 각 전수화 비교변수 X 값과 함수관계가 있는 변수으로써 표본조사 자료에는 포함되어 있으나 전수 자료에는 없는 변수이며, 이 글에서는 '전수화 대상변수'라고 하였다. O/D 및 P/A 전수화 과정에서는 통행목적별 통행생성량(trip production)과 O/D 및 P/A 쌍별 통행량이 전수화 대상변수 Y가 되는 것이다. 즉 전수 자료와 표본자료에 모두 포함되어 있는 전수화 비교변수 X 값을 이용하여 전수화 계수를 계산하고, 그 전수화 계수를 표본 Y 값에 곱함으로써 모집단 Y 값을 추정하는 것이 전수화 과정인 것이다.

하지만 위와 같은 가장 단순한 전수화 분석은 매우 까다로운 무작위성의 조건을 만족시킬 때에만 가능하다. 즉 앞의 예에서 모집단의 구성원 n의 가족 수, 차량보유대수, 거주주택형태 등과 같은 카테고리 분류의 분포가 표본 자료 구성원 k의 분포와 동일하여야 만이 완벽한 무작위 표본추출에 의한 자료라고 고려될 수 있을 것이다. 즉 표본 구성원의 분포가 모집단의 분포를 완전히 대표해야 한다는 조건을 만족시켜야 하는 것이다. 만일 이와 같은 표본이 모집단과 동일한 분포 또는 비율이 아닐 경우는 모집단의 분포와 맞추는 추가적 분석이 필요하게 된다. 또한 전수화 된 최종적 값을 알고자 하는 전수화 대상변수 Y 값에 대해서는 k번째 표본의 Y값 조사 과정에서 일부 값이 누락되었거나 조사 과정상의 오류가 없다는 조건을 만족시켜야 한다. 만일 조사 값의 누락

및 오류가 있을 경우에는 표본의 참 값(true value)을 가능한 옳게 추정하는 과정을 거친 후에 표본 자료 값을 수정할 필요가 있다. 즉 표본의 무작위성을 만족시키지 못하거나, 표본조사 자료 값 상의 오류가 있을 경우에는 표본의 대표성이 보장될 수가 없으므로 각 카테고리별로 전수화 계수 및 표본 조사 참값을 올바르게 추정하기 위해서는 복잡한 추가적인 분석과정을 거쳐야 한다. 그래야만 모집단의 진실 값에 가깝게 추정된 전수화 자료를 얻을 수 있을 것이다.

교통모형의 계수 값(parameter) 추정을 위한 모형정산(model calibration) 과정에서는 전수화 과정이 필요하지가 않다. 또한 표본조사 자료로부터 모집단의 총량적 값이 아닌 모집단의 비율(rate, share) 또는 평균값과 같은 값을 추정하기 위해서도 전수화 과정은 필요하지 않으며 다만 통계적 유의성 검증을 통해 모집단의 참값을 추정하면 된다. 전수화 과정이 필요한 경우는 모집단의 기술 통계치(descriptive statistics)로써 연구대상지역 전체 혹은 개별 교통존의 가구 수, 차량등록대수, 통행량 등과 같은 총량 값을 표본자료로부터 파악하고자 할 경우이다. 이와 같은 전수화 된 자료는 모집단 총량의 추정 값으로써 시계열적 변화 및 추세파악, 정책목표달성 측정치 산출, 그리고 무작위 표본자료(random sample)가 아닌 표본자료를 이용한 모형 정산 시의 가중치 계산(weighting) 등에 활용되고 있다. (Stopher and Jones, 2003) 교통체계분석 혹은 교통수요분석에서 표본자료의 전수화 과정이 필요한 항목은 교통존별 통행목적별 총 통행발생량, 통행목적별 기종점(O/D, origin-destination) 통행량 및 생성-유인(P/A, production-attraction) 통행량, 교통수단별 이용 통행수, 대중교통 역별 승하차 승객 수, 도로 지점별 차종별 단위시간 당 총 교통량 등의 자료일 것이다. 본 글에서는 교통존의 목적별 총 통행발생량과 교통존 쌍 간의 O/D 및 P/A 통행량을 추정하기 위한 전수화 과정에 초점을 두고 논의하고자 한다.

III. O/D 및 P/A 전수화의 정확성 확보 한계성

1. 표본 O/D 및 P/A 의 기본 특성으로 인한 한계성

가구당 혹은 개인별 통행발생모형, 개인 통행별 교통수단선택 모형, 노선

선택모형 등과 같은 통행행태모형의 정산(calibration)을 위해 필요로 하는 조사 자료는 비교적 작은 규모의 표본자료만을 필요로 하고 있다. 즉 교통관련 선택행태와 그 선택에 영향을 줄 것으로 보이는 요인들 간의 상관관계 원칙을 나타내는 모형들의 계수 값을 추정하고자 할 때 필요한 표본 자료의 수는 비교적 크지 않아도 높은 통계적 유의성 수준에서 모수 값을 추정할 수 있다. 이와는 달리 각 존쌍(zone pair) 간 표본 O/D 및 P/A 통행량은 각 개별표본의 통행들을 출발존-도착존 및 통행생성존-통행유인존의 존쌍으로 분류하여 모든 통행을 존쌍별로 합한 총량적 개념의 통행수이다. 따라서 표본 O/D 및 P/A 자료는 그 자체가 하나의 표본(sample unit)이 된다. 즉 표본 O/D 및 P/A 표가 표본에 포함된 모든 통행자들의 공간적 통행분산패턴을 나타내는 하나의 표본인 것이다. 그러므로 O/D 및 P/A 표본자료는 각 가구 또는 개인의 통행발생량 표본자료와 같이 여러 개의 표본을 이용하여 표본 평균과 표준편차를 계산할 수 있는 자료와는 근본적인 차이가 있다. 표본조사자료의 전체 합이 하나의 표본자료(sample unit)가 되는 O/D 및 P/A 표본자료는 평균 값 또는 분산 (혹은 표준편차)의 값을 알아 낼 수가 없다는 한계점이 있는 것이다. 따라서 통계학에서 잘 알려진 적절한 표본규모 산정 계산식을 O/D 및 P/A 조사를 위한 표본규모 계산에 적용할 수가 없다는 어려운 점이 있다. 예를 들면 10 % 표본율로 표본조사가 이루어진 경우 특정 교통존 쌍간의 O/D 또는 P/A 표본통행량이 123인 경우는 모집단은 1,230통행정도가 존재한다는 의미가 되는 것이다. 하지만 123통행은 모든 표본 통행자들의 통행 중 특정 교통존 쌍간에 이루어진 통행의 합으로써 하나의 표본 값이 되는 것이다. 그러므로 이 자료로부터 평균값과 분산 값을 알 수가 없으므로 통계학적 적정 표본규모의 계산은 가능하지가 않다. 또한 특정 존쌍 간의 표본 통행량이 0인 경우는 모집단의 통행량이 0에서 9통행 사이 값을 가질 것이라는 추론이 가능할 뿐이고 정확한 모집단의 통행수를 추정할 수는 없는 것이다. 모집단에서 0에서 9사이의 통행만 존재하는 존 쌍에 대해 정확한 통행량 추정을 하기 위해서는 표본율을 10%에서 더 높은 표본율로 높이는 방법 외에는 없다. 하지만 방대한 지역의 많은 교통존 수를 갖는 현실의 O/D 및 P/A 조사에서 통행이 매우 적은 존 쌍까지 정확하게 추정하기 위해 일괄적으로 표본율을 높여 막대

한 비용과 시간을 추가시킬 수가 없다. 따라서 O/D 및 P/A조사의 경우 전수조사가 아니기 때문에 발생하는 표본오차(Sampling Error)는 피할 수가 없는 것이므로 이와 같은 오차는 현실적 한계로써 받아들여야만 할 것이다. 다만 조사설계, 표본설계, 코딩과정, 전수화 계수산출 등 조사 및 분석 과정에서 발생할 수 있는 오류를 최소화하는 것이 전수화 분석 상 최선의 방법일 것이다.

2. 완벽한 무작위 표본추출의 한계성

각 교통존별로 거주인구 기준에 따라 동일한 표본을 기준으로 가구통행 실태조사가 이루어진다고 할지라도 각 교통존별로 유효 표본율이 다르게 되는 것이 보통이다. 더욱이 통행발생량에 영향을 주는 가구수, 차량보유대수 및 소득 등의 범주별로 구분해서 본다면 표본율을 동일하게 확보하는 것은 더욱 어려워 완벽한 무작위 표본 추출은 현실적으로 매우 어렵다. 이것은 표본 1개 자료가 모집단의 몇 개의 자료를 대표하는가에 대한 값(전수화 계수)이 개별적으로 또는 범주별로 차이가 있을 수 있다는 것을 의미하는 것이다. 즉 완벽한 무작위 표본이 아니라면 특성에 차이가 있는 표본별로 전수화 계수를 별도로 적용하여야 하는 것이 원칙일 것이다. 하지만 개별 표본의 대표성을 고려하여 하나하나의 개별 표본별로 전수화 계수를 적용하는 것은 현실적으로 불가능하다. 따라서 집합화된 카테고리 분류에 따라 별도의 전수화 계수를 적용하는 것이 현실적일 것이다. 합리적 숫자 이내의 범주별로 별도의 전수화 계수를 적용하는 것이 무작위 표본추출 한계성에 의한 오류를 최소화하는 현실적 방법인 것이다. 즉 교통존별 그리고 가구특성에 의한 범주별의 조합에 의한 카테고리별 전수화 계수를 아래의 식(3)과 식(4)와 같이 계산하는 방법이 현실적 차선택일 것이다.

$$f_c = \frac{\sum_{n \in c}^N X_n^p}{\sum_{k \in c}^K X_k^s} \quad (3)$$

$$Y_c^p = f_c \left\{ \sum_{k \in c} Y_k^s \right\} \quad (4)$$

여기서 c : 공간 또는 특성에 따라 통행패턴 유사 집단으로 구분한 카테고리 c

하지만 위와 같이 단순화 하였음에도 불구하고 교통존과 가구특성별 범주의 조합(combination)에 의한 카테고리의 수 규모가 매우 커지는 경우가 많다. 따라서 카테고리별 전수화 계수 값 계산을 위해 필요한 카테고리별 전수화 비교변수 X 를 전수자료와 표본자료 모두에서 확보하기는 매우 어려운 것이 현실이다. 이와 같은 이유에서 카테고리별로 단순히 전수화 계수를 계산할 수 없는 처지에 놓이게 된다. 즉 각 범주의 조합에 의한 카테고리가 아니라 범주별 집합적 자료만 제공되는 경우가 많아 카테고리별 전수화 계수 값 추정을 위해서는 간접적인 추정 방법이 필요하게 된다. 따라서 Stopher와 Stecher (1993) 그리고 김순관 (2005)에서 적용하였듯이 Furness 기법 (또는 2 dimensional balancing 혹은 row-and-column factoring으로 알려진 기법)과 같은 방법을 통해 안정적 전수화 계수 값이 추정될 때까지 반복 계산하여 대략적으로 전수화 계수를 유추하는 방법을 적용하는 수밖에 없는 것이다.

3. 오류 및 누락 없는 가구통행실태 조사자료 수집의 한계성

표본자료 1개는 모집단의 여러 개 자료를 대표하고 있는 것이므로 표본 자료상의 오류 및 누락된 정보는 전수화 계수에 의해 증폭되어 모집단 추정 값에 왜곡을 가져오게 된다. 따라서 표본자료는 분류된 범주의 모집단을 가장 잘 대표할 수 있게 모든 통행목적, 통행수단 등이 정확하게 기입 조사되어야 한다. 하지만 설문 응답자의 기억 부족, 설문응답 피로감, 부주의성 및 대리설문응답(proxy response), 사생활 노출회피 등으로 인한 누락통행(missing trip), 비응답 항목(non-response item), 잘못된 정보 등이 발생하게 된다. 하지만 잘못된 자료를 방대한 표본자료 속에서 찾는 것도 매우 어려우며, 설혹 찾았다 할지라도 진실 된 정보자료가 무엇인지를 판단하기도 어렵기 때문

에, 정확한 자료로 복원시키는 것이 매우 어려운 일이다. 물론 이와 같은 오류를 극소화하기 위해 반복적인 설문응답자 접촉(repeated contacts)으로 반복 확인 절차를 밟음으로써 자료의 질을 높일 수는 있겠지만 현실적으로 비용과 시간의 제약이 있게 된다. 따라서 이와 같은 오류의 가능성에 대해 인식하고, 정확성은 다소 떨어지더라도 기존 정부공식자료와 같은 거시적이고 총량적 검증 자료를 통해 오류를 최소화하기 위한 검증 과정을 수행할 필요가 있을 것이다.

4. 전수화 비교변수와 전수화 대상변수에 대한 기초자료의 부족

O/D 및 P/A 전수화를 위한 첫째 기본조건은 모집단과 유사한 존쌍간 통행분포패턴의 표본자료가 필수적이다. 즉 카테고리별 표본 발생통행량과 표본 O/D 및 P/A 통행량과 같은 전수화 대상변수의 표본 자료가 기본적으로 필요하다는 것이다. 두 번째 기본조건은 이와 같은 표본의 공간적 통행분포패턴을 모집단 규모로 확대하기 위해 통행특성이 유사한 카테고리별 전수화 계수 값 계산을 위한 전수화 비교변수(X) 값이 표본 자료와 전수 자료에 함께 존재하여야 한다는 것이다. 하지만 이와 같이 2가지 조건의 기초자료를 다 갖추고 있지 못한 경우가 현실적으로는 많이 있게 된다. 2가지 기본조건이 갖추어 있지 못할 경우는 가장 근간이 되는 기초자료가 없는 상태이므로 O/D 및 P/A의 공간적 분포패턴과 통행발생량 규모에 있어 정확성이 떨어진다는 것은 불가피한 현상이다. 따라서 이 경우에는 철도, 고속버스, 항공과 같은 지역 간 대중교통의 승하차 승객 수 및 O/D 승객 수 그리고 고속도로의 요금 정산소에서 요금 정산소까지의 차량 O/D 자료 등과 같이 전수 기록이 있는 정부의 공식적 기초 자료를 최대한 활용하여야 할 것이다. 또한 검사선(screen line) 및 폐쇄선(cordon line)에 의한 교통존 그룹 간의 총량적 왕래 통행량 검증, 그리고 노선배정(traffic assignment) 분석결과와 주요 도로 구간에서의 관측 교통량 자료와의 비교 검증 분석을 통해 전수화 O/D 및 P/A 통행량 패턴 추정의 오류를 최소화하는 노력이 필요하다.

IV. O/D 및 P/A 전수화 자료의 신뢰성 증대 방안

1. 기초 조사자료의 정확성 확보

O/D 및 P/A 전수화 자료의 신뢰성을 증대시키기 위해서는 전수화의 기초자료가 되는 가구통행실태조사의 정확성을 가능한 범위에서 높이는 것부터 시작되어야 할 것이다. 전수화 분석의 입력 자료가 되는 기초자료에서 오류가 발생할 경우 전수화 O/D 및 P/A 자료의 신뢰성 확보는 매우 어렵다. 따라서 조사자료의 질적 향상은 매우 근본적인 과제인 것이다. Stopher와 Jones (2003)는 교통조사자료의 질적 수준을 평가하기 위해 일련의 조사에서 전수화까지의 과정을 요약정리하고, 그 과정의 각 단계별로 이론적 측면, 현실적 실행측면에서 얼마나 충실한가에 따라 교통자료의 질적 수준을 평가할 수 있음을 서술하였다. 그 조사에서 전수화까지의 과정은 1) 조사의 목적 규명, 2) 사전조사와 평가, 3) 설문지 설계 및 조사방법 설정, 4) 표본 추출 설계, 5) 가구통행 자료수집 집행계획 구축, 6) 표준적 조사 외의 새로운 조사기법 활용 계획, 7) 표본조사 실행, 8) 자료 코딩, 9) 자료 분석 및 전수화의 9단계로 구분되어 있다.

O/D 및 P/A 전수화 자료 구축은 Stopher와 Jones (2003)가 제시한 과정에 따라 한 단계 한 단계에서 주어진 예산과 시간 제약 하에서 가능한 오류와 오차의 규모를 최소화 하고자 하는 노력에 의해 이루어져야 할 것이다. 즉 표본설계(sample design)와 조사 설계(survey design)가 세심하게 설계되고, 계획된 후 치밀하게 표본 조사가 수행되어야 하는 것이다. 조사설계와 표본설계 및 자료 수정에 대해서는 Pas (1995)와 Stopher and Jones (2003)에 상세히 설명되어 있다. 또한 조사 자료상의 누락정보 및 비응답자료 등에 의한 자료 손실을 최소화하기 위한 자료 복구에 세심한 노력 역시 필요하다. 그리고 완벽한 무작위 표본자료의 분포를 갖지 못한 경우에는 적절하게 구분된 카테고리별 전수화 계수 산정이 가능하도록 전수화 비교변수의 기초자료 및 정부공개 자료의 확보 또한 필요하다.

2. 표본 통행발생량의 대표성 검증과 수정 작업

O/D 및 P/A 통행량과 같이 공간적 상호 교통량 교류규모를 파악하기에 앞서 시스템 전체의 총 통행발생량의 규모가 정확하게 전수화 될 수 있는가를 검증하고 수정하는 방법이 먼저 수행되어야 할 것이다. 그러기 위해서는 표본으로 조사된 각 가구의 통행발생량이 조사되지 않은 다른 가정의 통행발생량을 적절히 대표하고 있는가를 확인하고 그 규모를 전수화하여 확대시키는 작업이 필요하다. 이와 같은 전수화를 위해 가족수, 소득수준, 연령 등 가구특성에 따라 유사한 통행발생량 패턴을 보이는 가구특성별 범주를 적절하게 구분하는 작업부터 하게 된다. 그리고 이와 같은 가구특성 변수간의 조합으로 카테고리를 설정하고, 해당 카테고리의 표본 자료가 카테고리 내 모집단의 통행발생량 패턴을 잘 반영한다는 가정 하에서 통행발생량 전수화 분석이 수행되는 것이다. 따라서 각 카테고리에 분포된 표본자료 분포 패턴이 모집단의 분포 패턴과 유사하여야 만이 모든 카테고리에 동일한 값의 전수화 계수를 적용할 수가 있다. 만일 그렇지 않을 경우에는 각 카테고리별로 표본 한 가정이 대표하는 모집단의 가정 수에 차이가 생기므로 그 전수화 계수도 카테고리별로 다른 값을 갖게 된다.

일반적으로 아무리 표본설계와 조사설계가 무작위 표본추출을 위해 잘 설정되었다 할지라도 모든 카테고리의 분포가 모집단과 유사하게 조사되기는 거의 불가능할 것이다. 따라서 분석자가 설정한 각 카테고리별로 적절한 통행발생량 전수화 계수를 별도로 구축하여야 하는 작업이 필요하다. 이와 같이 표본조사 자료가 모집단을 적절하게 대표하지 못할 경우에 카테고리별로 전수화 계수를 추정하는 방법이 Stopher and Stecher (1992)와 김순관(2005)에 의해 제안되었다.

〈표 1〉의 예제로 Stopher and Stecher (1992)와 김순관 (2005)에 의해 제안된 통행발생량 카테고리별 전수화 계산 방법의 개념을 설명하고자 한다. 이해를 돕기 위해 X를 가구 특성 중에 가족수 변수라 하고 X=1을 2인 가족, X=2를 3인 가족, X=3을 4인 가족이라고 가정하기로 한다. 그리고 Z를 소득 변수라 하고 Z=1을 저소득층, Z=2를 중산층, Z=3을 고소득층이라고 가정하기로 한다. X 범주와 Z 범주의 조합인 카테고리별 표본 수는

<표 1> 통행발생량 전수화를 위한 가구특성별 범주와 카테고리

X, Z 속성 범주	X_1	X_2	X_3	표본 Z 합계	모집단 Z 합계	Row factor
Z_1	n_{11}^s	n_{12}^s	n_{13}^s	$n_{z_1}^s$	$N_{z_1}^p$	$R_1 = \frac{N_{z_1}^p}{n_{z_1}^s}$
Z_2	n_{21}^s	n_{22}^s	n_{23}^s	$n_{z_2}^s$	$N_{z_2}^p$	$R_2 = \frac{N_{z_2}^p}{n_{z_2}^s}$
Z_3	n_{31}^s	n_{32}^s	n_{33}^s	$n_{z_3}^s$	$N_{z_3}^p$	$R_3 = \frac{N_{z_3}^p}{n_{z_3}^s}$
표본 X 합계	$n_{x_1}^s$	$n_{x_2}^s$	$n_{x_3}^s$	n^s		
모집단 X 합계	$N_{x_1}^p$	$N_{x_2}^p$	$N_{x_3}^p$			
Column factor	$C_1 = \frac{N_{x_1}^p}{n_{x_1}^s}$	$C_2 = \frac{N_{x_2}^p}{n_{x_2}^s}$	$C_3 = \frac{N_{x_3}^p}{n_{x_3}^s}$			

예로 n_{11}^s (X_1 범주이며 동시에 Z_1 범주에 속하는 표본 수)와 같이 표현하기로 한다. X 범주별 또는 Z 범주별 합계는 예로 $N_{x_1}^p$ (X_1 범주에 속하는 모집단 자료수)로 표현하기로 한다. 앞에서 서술하였듯이 특정 카테고리의 가구 통행발생량을 전수화하기 위해서는 카테고리의 한 가구가 모집단에서 몇 가구를 대표하는가를 알아야 한다. 하지만 일반적으로 모집단의 자료가 속성변수의 범주별 합계에 국한된 경우가 많고 카테고리별 모집단의 자료 수는 일반적으로 구하기가 힘들다. 따라서 X 또는 Z의 범주별 전수화 계수(표에서는 column factor 혹은 row factor)의 계산은 가능하나 카테고리별 전수화 계수는 직접적인 계산이 가능하지 않다. 따라서 Stopher and Stecher (1992)와 김순관 (2005)은 Furness 기법(row-and-column factoring 혹은 two dimensional balancing)을 이용하여 카테고리별 전수화 계수를 추정하는 방법을 적용하여 구하였다. 즉 column factor를 해당되는 column의 카테고리 모두에 곱하여 column 전수화 조건을 만족시키도록 한다. 그리고 나머지 모든 column에 대해서도 같은 방법으로 계산한다. 그러면 row의 전수화 조건이 맞지 않게 되므로 다시 row factor를 해당되는 row에 있는 모든 카테고리에 곱하여 다시 row 전수화 조건을 맞추도록 한다. 역시 모든 row에 대해서 같은 방법으로 계산한다. 이와 같은 계산을 row factor와 column factor 값이 1에 가깝게 수렴할 때까지 반복하여 계산함으로써

최종적인 카테고리별 전수화 계수를 추정하도록 하였다. 즉 I 번의 반복 계산을 통해 얻은 column 이 c 그리고 row가 r인 카테고리의 전수화 계수는 식(5)와 같이 계산이 된다.

$$gf_{cr} = C_c^{i=1} \cdot R_r^{i=1} \cdot C_c^{i=2} \cdot R_r^{i=2} \cdot \dots \cdot C_c^{i=I} \cdot R_r^{i=I} \quad (5)$$

따라서 카테고리 (c, r)에 속하는 표본 한 가구의 통행발생량에 gf_{cr} 을 곱하면 그 표본 자료가 대표하는 모집단 가구의 전수화 된 통행발생량이 산출되게 된다. 이와 같이 통행발생량 패턴이 유사한 카테고리별로 통행발생량을 전수화 함으로써 시스템 내 통행발생량의 총량적 규모를 비교적 정확하게 추정할 수 있다.

3. 표본 O/D 및 P/A 통행의 전수화 방법

가구통행실태 조사에서 얻은 통행목적별 표본 O/D와 P/A 통행량은 하나의 표본자료로써 통계학적 유의성 정도를 분석할 수가 없다. 따라서 O/D 및 P/A 자료의 신뢰성을 확보하기 위해서는 조사 과정에서 오류 및 누락된 통행이 없이 교통존 간의 교류 통행량 패턴을 가능한 정확하게 조사하여 전수화에 적용하는 수밖에 없을 것이다. 즉 조사된 O/D 및 P/A 패턴이 모집단의 패턴을 대표할 수 있게 잘 조사되었다는 전제 외에는 다른 통계적 검증 방법은 없다. 다만 교통존 별 출발통행량(Origin)과 도착통행량(Destination) 혹은 생성통행량(Production)과 유인통행량(Attraction) 총량을 앞에 절에서 설명하였듯이 가구 특성 카테고리별 통행발생량을 구하여 이 값을 이용하여 전수화 통행량을 추정하여 보정할 수는 있을 것이다. 좀 더 구체적으로 설명하면 전수화 계산된 존별의 출발통행량(Origin)과 도착통행량(Destination) 혹은 생성통행량(Production)과 유인통행량(Attraction)을 이용하여 표본 O/D 및 P/A와 함께 Fratar 모형 혹은 Furness 기법을 이용하여 전수화를 수행함으로써 전수화된 O/D 및 P/A의 존쌍별 통행량을 추정할 수가 있다. 이때 주의할 것은 표본 O/D와 P/A 속에 포함된 zero cell에 대한 처리를 어떻게 합리적으로 할 것인가 하는 문제이다. 즉 표본에

서 특정 O-D 쌍 또는 P-A 쌍에 통행량이 없다고 두 지점간의 통행이 없는 것을 의미하는 것이 아니기 때문이다. 즉 표본율을 증가시켰을 경우 관측 통행량이 있었을 텐데 표본율이 적어서 관측이 안 되었을 수도 있기 때문이다. 이와 같은 경우 0과 1 사이의 작은 값인 seed trip 값을 임의로 적용하거나, 인접 유사한 특성의 교통존 간 통행량을 적용하거나, 나중에 중력모형과 같이 통행분포모형에 의해 추정하는 방법이 있을 수가 있다. 어느 방법이 더 이론적으로 우수하다고 단정할 수가 없으므로 분석자의 판단에 따라 한 방법을 선택하여 추정하여야 할 것이다.

4. 링크 관측 교통량에 의한 검증과 수정

표본조사 자료를 기초로 한 O/D 및 P/A 전수화 과정이 완료되면, 전수화 O/D 및 P/A 추정 값을 관측 교통량과 비교하여 전수화 된 자료의 정확성을 검증하는 과정이 반드시 필요하다. 이와 같은 검증 방법 중의 하나가 검사선(screen line)과 폐쇄선(cordon line) 조사 자료와의 검증이다. 즉 검사선에 의해 크게 분할된 두 지역 간의 통행량 교류, 폐쇄선의 내부와 외부 간의 통행량 교류를 총량적 개념 차원에서 관측 교통량과 전수화 자료가 일치하는가를 검증하고 수정하는 과정이 필요하다. 이 검증방법은 두 지역 간의 통행 교류량에 대한 검증으로써 총량적 차원의 통행량에 대한 적합성을 판단하는 검증과정이다. 또 다른 검증 방법으로는 전수화 O/D 자료에 의한 노선배정(Traffic Assignment) 분석 결과와 관측 지점교통량을 비교하여 전수화 O/D 자료와 네트워크 자료 및 노선배정 모형의 정확성을 동시에 검증하는 방법이 있다. 이것은 통행량의 총량적 적합성과 함께 노선배정에 따른 교통량의 공간적 분산 배분정도의 적합성을 동시에 검증하는 과정이라고 볼 수 있다. Stopher and Meyburg (1975)에서도 간략하게 서술된 것으로 검사선 검증, 폐쇄선 교통량 검증, 노선배정 분석결과인 링크 교통량과 관측 지점교통량의 비교 검증 등을 통해 전수화 O/D를 수정하는 추가적인 방법인 것이다. 이와 같은 검증 분석과정을 적절하게 수행하기 위해서는 O/D 표본조사와 함께 검사선, 폐쇄선 및 지점교통량 등에 대해 관측 조사지점, 조사내용, 조사시점에 대해 체계적으로 계획되고 수행되어야

한다. 이와 같은 검증 과정의 분석이 적절하지 못할 경우 전수화 O/D 자료의 치명적 오류를 수정할 수 있는 마지막 기회가 상실되는 것과 마찬가지로 이로 검증 기초자료 구축과 검증 방법론 개발에도 많은 관심을 가져야 할 것이다.

5. 교통관련 자료의 통합적 활용과 현행화

표본자료의 확대 및 기초자료의 구축은 단시간에 이루어지지 않으며 또한 장래에도 예산과 시간의 제약으로 항상 정확성에는 한계성이 있게 마련이다. 따라서 이와 같은 현실적 제약성을 인정하면서 전수화 O/D 및 P/A 자료의 정확성을 효과적으로 높일 수 있는 방안을 찾을 필요가 있다. 비교적 같은 규모의 예산과 연구 기간을 가지고도 정확성을 높이기 위해서는 다양한 기관에서 그리고 여러 시점에서 조사되는 자료를 통합적으로 연계시켜 활용할 수 있는 체계를 구축할 필요가 있을 것이다. 즉 O/D 및 P/A 표본자료 구축의 표준화와 중앙DB관리체계를 갖추어서 자료 간의 호환성 확보를 통해 통합적 자료 활용 효율성을 극대화 할 필요가 있다. 과거와 현재 그리고 미래에 다양한 사업에 의해 조사되는 자료를 최대한 활용할 수 있는 통합적 자료 관리체계가 필요하다고 생각된다. 그 예로 자료 간의 호환성 (compatibility)을 확보하여 각각 다른 지역에서 다른 목적으로 조사된 자료일지라도 전국 O/D 전수화에 귀중한 기초 자료로써 사용될 수 있다고 보인다. 또한 조사년도의 전수화 O/D 및 P/A 자료를 매 년도마다 현행화 (up-date)하는 과정이 이루어질 수 있도록 제도화할 필요가 있을 것이다. 즉 매해 관측된 검증기초자료를 이용하여 검사선 및 폐쇄선에 대한 검증분석 (validation)을 수행하여 O/D 및 P/A를 가장 최근의 자료로 현행화함으로써 연차별로 지속해서 자료의 질적 수준을 높여가는 노력을 할 필요가 있을 것이다.

V. 맺는말

이론과 원칙에 충실하게 수행하여 얻은 표본조사, 모형정산 및 전수화 자

료를 활용하여 통행자들의 도시 및 지역 간 통행패턴을 이해하고, 통행행태 모형의 구축하는 것은 국가 및 지방정부의 과학적인 주요 교통정책 분석을 위해 가장 기초적이며 필수적인 연구이다. 특히 막대한 자원과 시간 그리고 방대한 표본자료를 요구하고 있는 O/D 및 P/A 표본조사와 전수화 작업은 많은 어려움과 노력이 투입됨에도 불구하고 그 업적이 교통관련 의사결정자에게는 가시화되지 못하여 그 중요성을 인정받지 못하는 경우가 많아 안타까운 일이다. 하지만 선진화 사회에 가면 갈수록 교통 분야에서 섬세성, 정확성, 예측기능의 향상을 정밀하게 요구되어지고 있으며, 이와 같은 정교한 기술응용은 지속적으로 성실하게 구축된 기초 연구를 바탕으로 이루어진다는 사실을 인식할 필요가 있다. 그러므로 통행자들의 정확한 통행행태패턴과 공간적 통행분산 패턴인 O/D와 P/A의 정확한 이해가 없이는 교통 분석의 정확한 예측 능력은 기대할 수가 없다는 점을 강조하고 싶다. 또한 이론과 원칙을 기초로 한 적절한 분석의 전 과정이 단계별로 충실히 이행될 때에만 질 높은 기초자료가 구축될 수 있을 뿐이고, 다른 지름길은 없다는 사실 또한 강조하고 싶다. 따라서 이론과 원칙에 기초하여 적절한 규모의 표본 수 확보가 가능하고, 조사 자료의 충분한 분석기간을 가질 수 있는 환경과 구조적 체계를 갖추어 반복적이며 지속적인 교통 기초자료의 구축이 될 수 있기를 바라는 마음이다.

참고문헌

1. Pas, Eric, Resource Paper for Survey Methodologies Workshop, Conference on Household Travel Surveys: New Concepts and Research Needs, TRB, March 12-15, 1995.
2. Stopher and Meyburg, Urban Transportation Modeling and Planning, Lexington Books, 1975.
3. Stopher and Stecher, Blow Up: Expanding a Complex Random Sample Travel Survey, Transportation Research Record, No. 1412, TRB, 1992.
4. Stopher and Jones, Developing Standards of Transport Survey Quality, Chapter 1 in Transport Survey Quality and Innovation edited by

Stopher and Jones, Pergamon, 2003.

5. 김순관, OD조사 전수화 방법, 교통기술과 정책, 제2권 제4호, 12월호, 2005.
6. 김익기, 교통수요분석에서 통행목적별 O-D 접근방법과 P-A 접근방법의 이론적 비교연구, 대한교통학회지, 제15권, 제1호, pp.45~62, 1997.
7. 김익기, 교통수요분석을 위한 정산과 예측 절차에 대한 재고찰, 교통기술과 정책, 제3권 제1호, 3월호, pp.87~106, 2006.