

OD조사 전수화 방법



김순관

1. 전수화 개념

- 전수화란 표본자료에 적정한 계수를 적용함으로써 전체 모집단의 특성과 최대한 유사하게 맞추는 일련의 과정으로, 전수화 과정에서 가장 중요한 부분은 표본에 적용할 전수화계수를 산출하는 것이다.
- 전수화계수란 조사된 표본하나가 대표할 수 있는 크기를 말하는데, 예를 들어 전수화계수가 50이란 것은 표본하나가 모집단에서는 50개를 대표한다는 것을 의미하는 것이다. 일반적인 전수화계수 산출식은 다음과 같다.

$$Exp_{ij} = \frac{Pop_{ij}}{Sam_{ij}}$$

여기서, Exp_{ij} : i 존, j 카테고리의 전수화계수

Pop_{ij} : i 존, j 카테고리의 모집단크기

Sam : i 존, j 카테고리의 표본크기

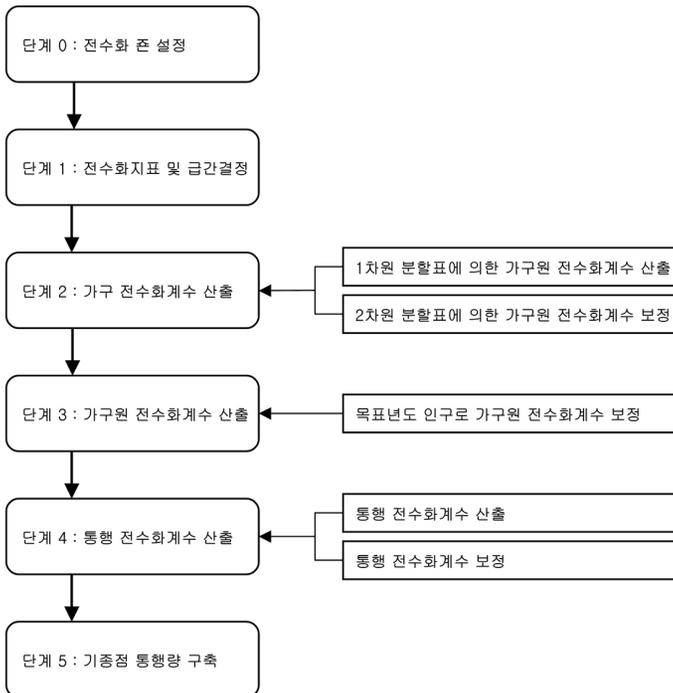
- 전수화계수 산출시 전체 조사지역을 대상으로 모집단크기와 표본크기를 이용하여 하나의 전수화계수를 산출하면 모든 지역에 동일한 전수화계수가 적용됨으로써 각 지역의 특성을 반영할 수 없는 문제점이 발

생하게 되는데, 이러한 문제점은 전수화계수를 각 지역별로 사회경제 지표별로 세분화하여 적용하면 어느 정도 개선할 수 있다.

- 전수화계수를 지역별로 사회경제지표별로 카테고리를 세분화하여 전수화계수를 산출하면 표본의 대표성이 좋아지나, 카테고리가 너무 세분화 될 경우에는 표본자체가 없는 카테고리(zero category)가 발생할 수 있어 과소추정의 위험이 있다. 따라서, 적절한 크기로 카테고리를 세분화하는 것이 중요하다.

II. 전수화 과정

- 전수화는 단계별로 순차적으로 진행되는데, <그림 1>은 단계별 전수화 과정을 도식화 한 것이다.



<그림 1> 전수화과정 흐름도

1. 단계 0 : 전수화 존 설정

- 조사된 모든 지역에 같은 전수화계수를 적용하면 각 지역의 특성을 반영할 수 없는 문제가 발생하고, 또 소존별로 각각 전수화계수를 적용하기에는 존별 표본수가 너무 적어 적정 표본수에 미달하는 존들이 있으므로 특성이 유사한 몇몇의 소존을 묶어서 전수화존을 설정한다.
- 전수화 존을 설정한 구체적인 방법은 다음과 같다.
 - 전수화존을 설정할 때에는 읍·면·동의 인구, 인구대비산업종사자수비율, 주거형태를 타나내는 공동주택·단독주택·기타주택의 거주비율, 경제척도를 판단할 수 있는 자가·전세·월세 비율 및 지리적 특성 등을 사용하여 전수화존으로 묶이는 소존들의 특성이 최대한 동질성을 띠게 한다.
 - 전술한 사회경제지표 카테고리와 지리적 특성에 의한 1차 분석결과를 바탕으로 전수화존을 설정한 결과 각 전수화존의 인구가 평균치보다 너무 많거나 적을 경우 해당 전수화존을 분할하거나 성격이 가장 유사한 인접 전수화존으로 병합한다.

2. 단계 1 : 전수화지표 및 급간 결정

- 전수화지표란 표본자료로 모집단을 추정할 때 기준으로 사용하는 변수로써 사용되는 변수는 표본과 모집단에 공통으로 존재하여야 한다.
- 각 계층별 통행특성을 충분히 반영하는 대표변수를 찾기 위하여, 통행특성과 상관관계가 높은 여러 가지 사회경제지표를 검토하여 그 중 표본과 모집단에 공통으로 존재하는 변수를 전수화지표로 설정한다.
- OD전수화를 위해 사용되는 모집단 자료는 통계청에서 실시하는 「인구주택총조사」의 가구자료이며, 기존 OD조사의 조사항목과 「인구주택총조사」의 조사항목에 공통으로 조사되는 변수는 “가구원수”, “미취학아동수”, “주택점유형태”이다. 따라서 이들 세가지 항목이 전수화지표로 사용될 수 있다.
- 선정된 전수화지표의 급간은 가구당 통행수를 종속변수로 각 전수화지

표를 독립변수로 하는 일원분산분석(One-Way ANOVA)을 수행하고, 각 급간끼리 평균을 비교하는 다중비교(Multiple Comparison)을 통해 결정한다.

- '2002 서울시 가구통행실태조사'에서 최종적으로 설정된 전수화 지표별 급간은 다음과 같다.

〈표 1〉 '2002 서울시 가구통행실태조사' 전수화 지표별 급간

전수화지표	급간
가구원수	1인, 2인, 3인, 4인, 5인 이상
미취학아동수	0인, 1인, 2인 이상
주택소유형태	자가, 전세 및 월세

3. 단계 2 : 가구 전수화계수 산출

- 가구 전수화계수는 전수화존별, 가구원규모별, 미취학아동수별, 주택점유형태별로 각각 산출하는 것이 가장 좋은 방법이나, 전수화존, 가구원규모, 미취학아동수, 주택소유형태의 급간은 '2002 서울시 가구통행실태조사'의 경우 각각 296개, 5개, 3개, 2개의 카테고리를 가지고 있기 때문에 전수화 계수를 산출해야 할 총 카테고리수는 8,880이며, 각 전수화 존별로는 30개가 된다. 따라서, 각 전수화존별 30개의 사회경제지표 카테고리를 그대로 유지시키면서 전수화계수를 산출하면 표본이 없는 카테고리가(zero category)발생하게 되고, 결국 과소추정의 문제가 발생한다.
- 이러한 문제점을 해결하기 위하여, 전수화계수 산출시 설정된 3개의 전수화지표를 3차원으로 분할하여 한번에 적용하는 대신, 1개지표의 1차원 분할표 또는 2개지표를 조합한 2차원 분할표를 순차적으로 적용하여 전수화계수를 산출하고, 산출된 전수화계수가 수렴할 때까지 일련의 과정을 반복하는 Furness 방법을 사용할 수 있다.

4. 단계 3 : 가구원 전수화계수 산출

- 조사된 각 가구원에 전 단계에서 산출된 가구 전수화계수를 적용하여

가구원별 전수화계수를 산출하는 과정으로 산출된 가구원 전수화계수는 전수화존별 전수화지표별로 동일한 전수화계수를 가지게 된다.

- OD조사와 인구주택총조사의 조사년도가 일치하지 않을 경우 사회경제지표 현황을 제대로 반영하지 못하게 된다. 또한, 전수화존을 기준으로 전수화지표가 산출되기 때문에 전수화존 내에 속하는 소존간 사회경제지표의 편차는 반영하지 못하게 된다.
- 따라서, 전수화존별 가구원 전수화계수를 전수화존내 소존별 사회경제지표를 기준으로 보정하는 과정이 필요하며, OD조사년도의 사회경제지표를 반영하는 과정이 병행되어야 한다. '2002 서울시 가구통행실태조사'의 경우 사회경제지표중 통행량에 직접적인 영향을 주고 확보 가능한 것은 인구이기 때문에 목표년도 사회경제지표 반영은 인구지표로 한정하였으며, 존간 사회경제지표 편차 역시 인구로 한정하여 반영하였다.

5. 단계 4 : 통행 전수화계수 산출

1) 단계 4-1 : 통행 전수화계수 산출

- 가구원 전수화계수를 통행인구별 각 통행자료에 적용하여 통행 전수화계수를 산출하는 과정이다.
- 가구원 전수화계수를 단순히 통행자료에 적용하여 산출한 통행 전수화계수는 통행목적별 발생율의 차이를 반영하지 못하고, 전수화지표 카테고리별로 모든 통행의 목적이 같은 전수화 계수를 가지게 된다. 따라서, 통행목적별 특성의 차이를 최대한 반영하기 위하여, 전수화존내 동별·가구원수별·통행목적별 전수화계수가 동일하게 적용될 수 있도록 전수화계수를 조정하는 과정이며, 이때 전수화존내 목적별 전수화계수 총량이 변하지 않게 고정한다.

2) 단계 4-2 : 통행 전수화계수 보정

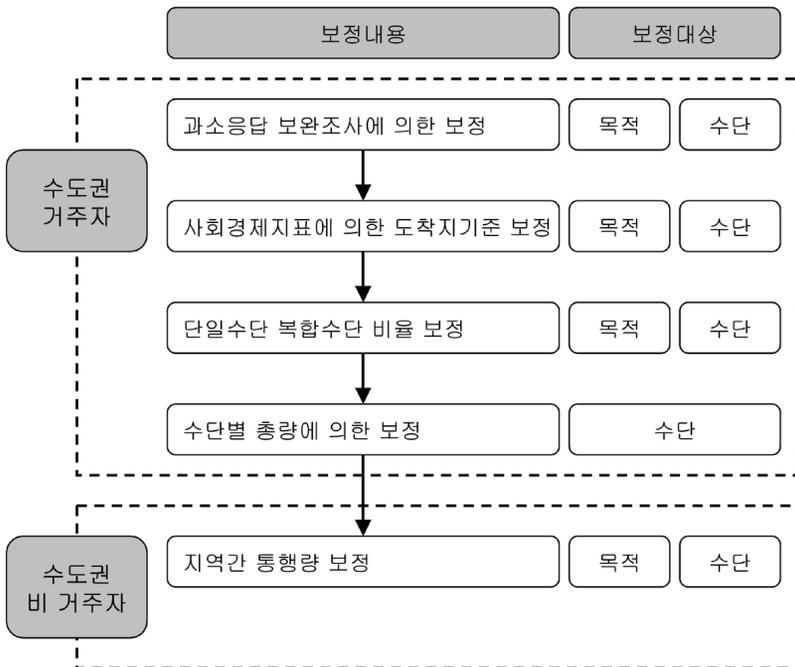
- 가구통행실태설문조사는 일반적으로 조사의 특성상 비정규적인 통행(단거리통행, 심야통행, 친교, 위락, 쇼핑, 개인용무등)을 비롯하여,

특정수단(택시, 오토바이, 화물차등)의 통행이 누락되기 쉽다. 또한, 가구통행실태조사는 조사대상지역내에 거주하는 가구에 대해서만 하루동안 활동한 가구원의 통행일지를 기록하는 설문조사 형태로 이루어지기 때문에 비거주자가 조사대상지역내에서 활동하는 통행은 누락되게 된다.

- 따라서, 앞 단계까지 전수화된 통행량에 누락된 통행을 산출하여 추가하는 과정이 필요하다. 이렇게 누락된 통행을 보정하는 단계는 크게 다음의 5가지 단계로 구분할 수 있다.
- 가구통행실태조사의 특성상 하루의 통행이 완료된 후 작성되기 때문에 통행자가 모든 통행에 대해 기억을 못하는 까닭으로 불규칙적인 목적의 통행(단거리통행, 심야통행, 친교, 위락, 쇼핑, 개인용무)을 비롯하여 전반적으로 통행이 누락되는 경향이 있는데, 이렇게 누락된 목적통행은 과소추정의 원인이 된다. 따라서, 가구통행실태조사 설문지상에서 누락된 통행을 보정하기 위한 첫 번째 단계는 지역별로 가구의 특성이 동일한 집단에서 통행의 누락을 최소화시키기 위하여 별도의 조사방법으로 조사된 자료를 이용하는 것이다. 이때, 목적통행을 보정함에 있어서 목적과 연계하여 수단통행도 함께 보정하여야 한다.
- 1단계 보정에 사용된 과소응답보완조사도 여전히 통행의 누락이 있을 수 있고, 조사대상이 아닌 수도권 비거주자에 의한 통행부분을 보정하기 위하여 도착지별 유인(Attraction)요소를 고려한 도착지 기준 보정이 두 번째 단계이다. 예를 들어, 도착지 기준의 사회경제지표(업종별 근로자수, 수용 학생수, 인구수 등)를 독립변수로 하는 통행목적별 도착통행량 선형회귀 모형을 구축한 다음, 이를 이용하여 회귀식 상단부와 하단부의 극단치 통행량을 모형값으로 대체하는 방법을 적용할 수 있다.
- 셋째, 대중교통수단중 타 수단에 접근하기 위한 혹은 환승을 위한 수단은 설문조사에서 누락되는 경향이 많기 때문에, 이를 보정하기 위하여 수단간의 단일수단과 복합수단의 구성비를 별도로 조사된 조사자료를 이용하여 보정하는 단계이다.
- 넷째, 승용차, 대중교통 등 일상적으로 이용되는 수단은 설문조사시

누락되는 경우가 적으나, 상대적으로 이용량이 적거나 교통수단간 환승시 접근수단으로 이용되는 교통수단(택시, 마을버스, 이륜차 등)이 누락되는 경우는 많기 때문에 이를 보정하는 단계이다. 이는 별도로 조사되었거나, 공식적으로 집계되어 있는 수단별 총량집계와 전수화된 수단별 통행량을 비교하여 보정할 수 있다.

- 다섯째, 수도권 비거주자에 의한 통행 보정은 수도권 거주자 대상인 가구통행실태조사 자료만으로는 한계가 있기 때문에 별도의 방법으로 보정하는 단계이다. 예를 들어, 별도로 실시한 보완조사자료(터미널 조사, 코든라인 교통량조사) 및 고속도로·국도의 교통량 자료를 이용하여 수도권과 수도권 외부 전국지역간의 통행량을 보정할 수 있다.
- <그림 2>는 통행 전수화계수 보정과정을 도식화한 것이다.



<그림 2> 통행 전수화계수 보정과정