

가구통행실태조사의 표본을 산출방법



김채만



류시균



구정은

1. 서론

「교통체계효율화법(법률 제8050호)」에 의하여 매 5년마다 실시되고 있는 가구통행실태조사는 국가 및 지방자치단체 교통정책 수립의 근거로 사용되고 있는 중요한 조사로서 방대한 예산이 소요되는 대규모의 사업이다. 수도권에서는 2006년 약35억의 예산을 투입하여 경기, 서울, 인천이 공동으로 가구통행실태조사를 수행하고 있으며, 2007년 조사결과를 발표할 예정이다. 또한, 동시에 5개 광역권에 대한 가구통행실태조사가 건설교통부 주관하에 수행되고 있다.

가구통행실태조사는 표본 조사로서 전수화 과정을 거쳐 조사대상지역의 기·종점자료 산출을 목표로 한다. 산출된 기·종점자료는 국가 SOC사업의 우선순위와 민자사업의 요금결정의 근거로 자료로 사용될 뿐만 아니라 지방자치단체의 법정계획인 도시기본계획, 교통정비기본계획, 도로정비기본계획, 지방대중교통계획 등을 수립하는 기초자료로 사용되고 있다. SOC사업의 우선순위와 민자사업의 요금결정을 위해서는 구축된 기·종점자료의 정도(精度)가 명확히 정의되어 있어야 한다. 자료의 정도(精度)는 분석

김채만 : 경기개발연구원 교통정책연구부, cmkim@gri.re.kr, 직장전화:031-250-3157, 직장팩스:031-205-3116
 류시균 : 경기개발연구원 교통정책연구부, ryu@gri.re.kr, 직장전화:031-250-3282, 직장팩스:031-205-3116
 구정은 : 경기개발연구원 교통정책연구부, ingodslove@gri.re.kr, 직장전화:031-250-3157, 직장팩스:031-205-3116

결과에 대한 이해집단인 중앙정부, 지방정부, 민간투자기관, 시민의 사회적 합의를 이끌어 내는 중요한 요소이다.

가구통행실태조사는 표본 조사이기 때문에 자료의 정도(精度)는 표본율에 의해서 결정되어 진다. 우리나라의 가구통행실태조사는 「교통조사지침」에 의해 자료의 정도(±30%)를 규정하고 있으며, 이를 확보하기 위한 최소 표본율을 제시하고 있다. 그러나, 교통조사지침에 규정된 정도(精度)가 무엇을 추정하기 위한 정도인지 불명확하게 정의되어 있다. 따라서, 본 연구에서는 교통전문가들이 일반적으로 인식하고 있는 통행수단별 기·종점통행량의 분포추정을 목표로 했을 경우의 일정한 정도를 확보하기 위한 표본율은 얼마인지 분석하고, 현재의 교통조사지침에서 제시하고 있는 표본율의 문제점과 개선방안을 제시하고자 한다.

II. 표본율 산정 방법

1. 교통조사지침

교통체계효율화법에 정의된 교통조사에 대한 교통조사 기준 및 방법을 「교통조사지침(건설교통부 고시 제2006-415호)」에서 규정하고 있다. 교통조사지침은 “공통적인 적용 기준”과 “조사부문별 적용기준”으로 구분하여 구성되어 있다. “공통적인 적용기준”에서는 표본 설계시 표본율은 첫째 추정하고자 하는 대상(평균의 추정, 총계추정, 모비율 추정, OD추정), 둘째 추정하고자 하는 대상의 신뢰수준, 셋째 목표 정도(精度), 넷째 모집단의 크기에 따라 결정되어야 한다고 규정하고 있다. “조사부문별 적용기준”에서는 개인통행실태조사의 최소 표본율은 식(1)과 같이 규정하고 있으며, 존의 인구규모에 따라 <표 1>를 병행 사용하도록 규정하고 있다.

교통조사지침의 개인통행실태조사부분 최소 표본율 산출은 논리적으로 몇가지 문제점을 갖고 있다. 첫째, 추정대상이 명확하게 제시되어 있지 않다. 둘째, 존의 인구에 따른 고정적인 최소 표본율은 교통존의 인구규모만 고려한 것으로 논리적으로 불합리하다.

$$\epsilon = z_{(\alpha/2)} \sqrt{(l-1) \cdot \frac{1-r}{r} \cdot \frac{1}{N}} \quad , \quad r = \frac{1}{[(\epsilon/z)^2 \times \frac{N}{l-1}] + 1} \quad (1)$$

- 여기서 ϵ : 상대허용오차(d/P)
 $z_{\alpha/2}$: 유의수준 α 에서의 신뢰계수
 l : 존 수
 r : 표본율(n/N)
 n : 표본의 크기(샘플의 인구)
 N : 조사대상 전체 인구수(모집단 인구)
 d : 절대오차
 P : 표본의 특성 비율

〈표 1〉 개인통행실태조사 표본율(교통조사지침)

교통존별 인구수	최소 유효표본율
인구 5천명 미만	3.6%
인구 5천명 ~ 1만명 미만	2.9%
인구 1만명 이상	2.4%

첫 번째 문제는 추정대상에 따라 최소표본율이 달라지고, 자료 이용자와 정책결정자에게 정확한 정보를 전달하지 못한다는 단점이 있다. 개인통행실태조사의 최소 표본율 산출식에서는 추정하고자 하는 대상에 따라 모집단의 크기와 표본의 특성비율이 달라진다. 따라서, 추정하고자 하는 대상에 따라 산출된 최소 표본율이 결정된다. 또한, 전수화된 자료를 이용하는 사용자측면, 분석결과를 활용하는 정책결정자 입장에서도 자료가 갖고 있는 특정측면에서의 오차를 인식하고 있어야만 사회적 합의를 이끌어 내는데 유리하다.

두 번째 문제는 비논리적이다. 교통존별 인구규모(모집단의 크기 : N)에서 따라 표본율을 차등적으로 적용하여야 한다는 점에서 논리적으로 타당하다. 그러나 표본율은 교통존의 인구규모 뿐만 아니라 다른 여러 가지 변수(정도(精度), 존수, 모집단의 크기, 유의수준 등)에 의해서 결정된다. 따라서, 〈표 1〉의 표본율은 목표정도, 존수, 모집단의 크기, 유의수준 등이 결정된 상태에서의 표본율 예시로 제시하는 것이 바람직하고, 일반적인 기준으로 적용하는 것은 비논리적이다.

2. 일본의 표본을 산정방법

일본 동경권은 1968년 이후 매 10년 간격으로 개인통행실태조사를 수행하고 있다. 가장 최근 조사는 1998년 4회차 조사이다. 동경권의 4회차 개인통행실태조사에 사용한 표본을 산출식은 일반통계이론으로부터 추정되었다.

일반 통계이론에서 모비율의 구간추정은 모집단의 크기(N), 표본수(n), 추정하고자 하는 모집단의 특성비율(p), 표본분포의 특성비율(P), 표본분포의 분산($\frac{N-n}{N-1} \cdot \frac{P(1-P)}{n}$)에 의하여 추정된다.

모비율 추정 구간의 편측간격을 나타내는 $z_{\alpha/2} \sqrt{\frac{N-n}{N-1} \cdot \frac{P(1-P)}{n}}$ 는 목표 허용오차(d) 보다 작거나 같아야 한다. 따라서 목표 절대오차를 유지하는 최소 표본의 크기는 식(2)과 같다.

$$d = z_{\alpha/2} \sqrt{\frac{N-n}{N-1} \cdot \frac{P(1-P)}{n}} \quad (2)$$

최소표본 추출율($r = n/N$)을 도입하여 식(2)를 정리하면 식(3)과 같다.

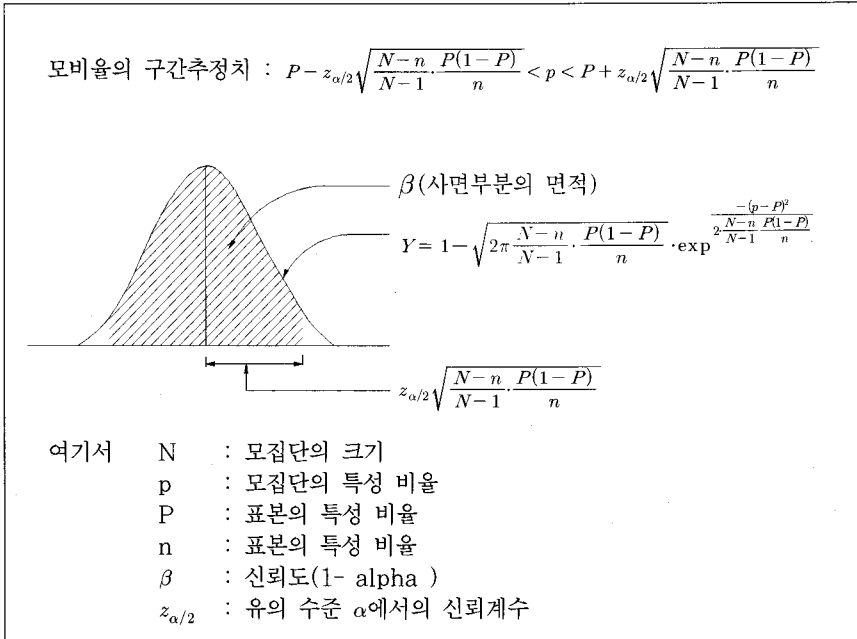
$$d = z_{\alpha/2} \sqrt{\frac{1}{N-1} \cdot \frac{1-r}{r} \cdot P(1-P)} \quad (3)$$

모집단의 특성 비율(P)을 반영한 상대적 오차($\epsilon = \frac{d}{P}$)를 도입하여 식(3)을 다시 정리 하면 식(4)와 같다.

$$\epsilon = \frac{d}{P} = z_{\alpha/2} \sqrt{\frac{1}{N-1} \cdot \frac{1-r}{r} \cdot \frac{(1-P)}{P}} \quad (4)$$

- 여기서 ϵ : 상대허용오차(d/P)
- $z_{\alpha/2}$: 유의수준 α 에서의 신뢰계수
- r : 표본율(n/N)
- n : 표본의 크기
- N : 모집단의 크기

- d : 절대오차
- P : 표본의 특성 비율



〈그림 1〉 모비율의 구간추정

1998년 4회 동경권 개인통행실태조사의 최소 표본율은 위 식을 이용하여 산출하였다. 산출 기준은 추정대상을 “존별통행목적별수단분담율”로 설정하고, 신뢰수준을 95%($k=1.96$), 조사정도(상대오차)를 20%로 설정하였다. 도시권전체의 인구 3,273만명, 5개 통행목적(대표통행목적), 4개

〈표 2〉 동경도시권의 표본을 산정결과(제3회 동경도시권 PT조사, 1998년)

신뢰 계수 (k)	상대 허용 오차 (F)	5세이상 인구 (인)[A]		목적 종류 수 (B)	존 수 (C)		모집단 크기1) (N=A*2.47/B/C)		특성비율 (P)	표본율(%)	
		도시권 전역	동경 구부		도시권 전역	동경 구부	도시권 전역	동경 구부		도시권 전역	동경 구부
1.96	20%	32,731,383	7,719,169	5	1,644	264	9,835	14,444	0.25 = 100/4수단	2.85	1.96

주 1) : 2.47은 1988년 동경권 조사시 인당 목적통행발생 원단위임

교통수단(대표교통수단), 1,644개 존수(동경구부 264개 존)을 기준으로 하여 동경권 존별통행목적별수단분담율을 추정을 위한 최소 표본율은 동경도시권 전역은 2.85%, 동경구부는 1.96%인 것으로 산정하였다.

3. 교통조사지침의 재해석

일반 통계이론과 일본 동경권 개인통행실태조사 표본을 산정방법을 참고하여 「교통조사지침」의 개인통행실태조사의 최소 표본율 산정식을 재해석해보면, “모비율 추정에 필요한 최소 표본크기 결정”을 하는 일반식인 식(4)의 변형임을 알 수 있다. 구체적으로는 추정대상은 “존별 인구의 지역간 통행분포 비율”이며, 모집단은 조사대상지역의 전체인구(통계이론상 조사대상 지역 전체인구를 존수로 나눈 존별 평균인구가 적합함)로 추정된다.

모집단의 인구수가 크면 $\frac{1}{N-1} \approx \frac{1}{N}$ 이고, 존별 인구가 모든 존으로 통행하는 비율이 동일하다고 가정시 $P=1/l$ (존수)⁶⁾이다. 이를 일본 동경권의 개인통행실태조사 표본율 산출 식인 식(4)에 대입($\frac{1}{N-1}$ 은 $\frac{1}{N}$ 로, $\frac{(1-P)}{P}$ 을 $(l-1)$ 로 대체)하면 교통조사 지침에서 제시하고 있는 식(1)이 유도됨을 알 수 있다.

$$\epsilon = \frac{d}{P} = z_{\alpha/2} \sqrt{\frac{1}{N-1} \cdot \frac{1-r}{r} \cdot \frac{(1-P)}{P}} = z_{\alpha/2} \sqrt{(l-1) \cdot \frac{1-r}{r} \cdot \frac{1}{N}}$$

교통조사지침에서 제시하는 개인통행실태조사 부문의 식을 재해석한 결과 다음의 사항들을 확인할 수 있다. 첫째 교통조사 지침에서 제시하고 있는 최소 표본율 산출 식은 일반식이 아니며 식(4)로 교체하는 것이 바람직하다. 둘째 조사를 통하여 추정하고자 하는 대상이 명시되어 있지 않아(존별 인구의 지역간 통행분포 비율로 추정됨) 이를 명확히 규정할 필요가 있다. 셋째 추정대상을 “존별 인구의 지역간 통행분포”를 대상으로 할 경우에

6) $P=1/l$ (존수) 적용시 $\frac{(1-P)}{P} = \frac{(1-\frac{1}{l})}{\frac{1}{l}} = \frac{l-1}{1} = (l-1)$

도 모집단의 크기(N)를 “조사대상 전체의 인구수”라고 정의한 것은 잘못된 것으로 “교통존별 인구(가구) 또는 교통존의 평균인구(가구)”로 정의하는 것이 바람직하다.

「조사지침」에 근거하여 2006년 수도권 가구통행실태조사 대상에 대하여 존 수(t)는 1,358개, 모집단 크기(N)는 조사대상전체의 가구기준(현재지침)과 존별 가구기준으로 각각의 경우에 대하여 최소 표본율을 산정하였다. 모집단의 크기를 존별 가구수 기준으로 표본을 산정결과 “존간 인구의 지역간 통행분포 비율”을 신뢰도 95%의 상대허용오차 30%로 추정하기 위해서는 최소 표본율이 91.26%이며, 모집단의 크기를 전체 가구수 기준으로 할 경우 0.76%인 것으로 추정되었다.

존별 가구수를 모집단으로 할 경우 최소 표본율 91.26%는 Smith, M.E(1979)의 연구와 J.d.e Ortuzar et al(2001) 연구에서 기·종점 분포 추정을 목표로 할 경우 존수가 증가하면 95% 신뢰수준의 20%이하 상대허용오차를 확보하기 위한 표본의 크기는 모집단 크기와 유사하다고 도출된 결과와 일치하는 것이다.

〈표 3〉 2006년 수도권 조사의 최소표본율 산정결과(현재 교통조사지침)

신뢰계수 ($z_{\alpha/2}$)	상대허용오차	표본율(%)	
		N (조사대상지역 전체 가구수)	N (존의 평균 가구수)
1.28 (90% 신뢰수준)	30%	0.33	81.67.
	20%	0.73	90.93
	10%	2.87	97.57
1.64 (90% 신뢰수준)	30%	0.54	87.97
	20%	1.20	94.27
	10%	4.62	98.50
1.96 (95% 신뢰수준)	30%	0.76	91.26
	20%	1.70	95.92
	10%	6.47	98.95
2.57 (99% 신뢰수준)	30%	1.31	94.73
	20%	2.89	97.58
	10%	10.64	99.39

Ⅳ. 표본을 산정 개선방안

가구통행실태조사의 최소 표본율은 기·종점자료의 정도((精度)를 결정하는 중요한 요소이다. 따라서, 「교통조사지침」의 표본율 산정 식은 모비율 추정에 필요한 표본크기 결정 식인 식(4)로 대체되어야 한다. 또한, 최소 표본율 추정을 위해 추정대상을 총 목적통행의 수단분담율로 하고 상대허용 오차를 ±20%로 하는 것이 바람직하다.

교통존의 인구규모별 고정된 최소 표본율 표(교통조사지침 P28)는 “조사 자료의 신뢰도를 향상시키기 위해서 최소 표본율 산출식에 적용되는 모집단의 크기는 준별 평균치를 사용하는 것보다는 개별 교통존의 모집단 크기를 사용하는 것이 바람직하다.”로 대체하여야 한다.

《교통조사지침의 수정》

〈 수정항 : “4.개인통행조사 다.”항 p27~28)

다. 조사대상 및 방법

- 조사대상은 ~.(수정사항 없음)
- 조사방법은 ~.(수정사항 없음)
- 개인통행실태조사의 최소 표본율 추정대상은 존에서 발생하는 총목적통행의 수단분담율 추정으로 하고(통행수단은 5개로 가정함), 신뢰수준 95%에서 상대허용 오차 ±20%가 유지되도록 다음 식을 이용하여 산출한다.

$$\epsilon = \frac{d}{P} = z_{\alpha/2} \sqrt{\frac{1}{N-1} \frac{1-r}{r} \frac{P(1-P)}{P}}$$

여기서 ϵ : 상대허용오차(d/P)

$z_{\alpha/2}$: 유의수준 α 에서의 신뢰계수

r : 표본율(n/N)

n : 표본의 크기

N : 모집단의 크기

d : 절대오차

P : 표본의 특성 비율

- 조사자료의 신뢰도를 향상시키기 위해서 모집단의 크기는 준별 평균치를 사용하는 것보다는 개별 교통존의 모집단 크기를 사용하여 교통존별 유효 표본율을 산출하는 것이 바람직하다(즉 교통존별로 유효 표본율을 차등 적용하는 것이 바람직하다).

본 연구에서 제안한 내용이 수정할 경우 2006년 수도권의 가구통행 실태 조사의 최소 표본율을 5가지의 추정대상을 가정하여 산출하였다. 산출결과 신뢰수준 95%에 상대오차 $\pm 20\%$ 의 존간 통행분포 추정을 위해서는 79~95%의 최소 표본율이 필요하고, 총목적통행의 수단분담율 추정에는 1.68%, 통행목적별 수단분담율 추정에는 7.67%의 표본율이 필요한 것으로 분석되었다.

산출결과 2006년 수도권가구통행실태조사는 최소 유효표본율을 2.0%로 가정하고 있으며, 이는 추정대상을 “총목적통행의 수단분담율”로 하는 것이 가장 유사한 것으로 분석되었다.

<전제조건>

- 인구수 : 22,692,457
- 존 수 : 1,358
- 원단위(목적통행량/인) : 2.06(2002 서울시가구통행실태조사 분석결과)
- 통행목적 수 : 5(출근, 등교, 귀가, 업무, 기타로 단순화 가정)
- 통행수단 수 : 5(승용차, 택시, 버스, 지하철, 기타로 단순화 가정)

<표 4> 2006년 수도권의 추정대상별 최소 표본율 산출결과(교통조사지침 수정시)

추정대상	N (모집단의 크기)	P (샘플의비율 (가정))	신뢰 수준 (%)	최소 표본율(%)	
				상대오차 30%	상대오차 20%
총목적통행의 존간 통행분포	인구수×원단위/존 수 = 34,686	1/존수 = 0.00074	95%	62.58	79.00
총목적통행의 수단분담율	인구수×원단위/존 수 = 34,686	1/수단수 = 0.2	95%	0.73	1.63
총목적통행의 존간 수단별 분포	인구수×원단위/존 수 = 34,686	1/(존수×수단수) = 0.00015	95%	89.31	94.95
통행목적별 존간 분포	인구수×원단위/(존수 ×통행목적수) = 6,937	1/존수 = 0.00074	95%	89.32	94.95
통행목적별 수단분담율	인구수×원단위/(존수 ×통행목적수) = 6,937	1/수단수 = 0.2	95%	3.56	7.67

V. 결론

국가 및 지방자치 단체의 여러 교통정책 결정의 중요한 근거자료인 통행 실태조사자료의 신뢰도는 교통정책의 신뢰도와 교통정책의 성패에 매우 중요한 영향을 미치는 변수이다. 또한 조사 자료의 신뢰도를 결정하는 가장 중요한 요소는 최소 표본율이며 최소 표본율은 추정하고자 하는 대상, 추정 대상의 신뢰수준 및 상대허용오차에 의하여 결정된다. 중앙정부에서 제정한 「교통조사지침」에서 최소 유효표본율을 제시하고 있으나 제시된 최소 유효 표본율 산정식은 추정대상이 불명확하고 이론적 근거가 미약한 비일반화된 식으로 혼란을 야기하고 있다. 따라서 본 연구에서는 통계이론에 기반을 둔 명확하고 일반화된 식과 문구로 교통조사지침을 개정할 것을 제안하였다. 교통조사의 정확성 및 일관성을 확보하기 위하여 본고에서 제시한 근거에 기초한 교통조사지침의 조속한 개정이 필요하다.

참고문헌

1. 동경도시권교통계획협의회(2004), “「동경도시권 가구통행실태조사」자료 이용 안내서”.
2. 김채만(2006), 경기도 기종점 자료의 신뢰도 증진방안에 관한 연구, 경기개발연구원.
3. Michael E. Smith(1979), “Design of small-Sample Home-Interview Travel Surveys”.
4. J.de Ortuzar et al(2001), “Modelling Transport -Third Edition”, Wiley.