

■ 論 文 ■

예산제약하에서 O/D 추정을 위한 최소표본을 결정

Sample Size Determination for O/D Estimation under Budget Constraint

신 희 철

(한국교통연구원 국가교통DB센터 책임연구원)

이 향 숙

(한국교통연구원 국가교통DB센터 연구원)

목 차

- | | |
|---|--|
| <ul style="list-style-type: none"> I. 서론 II. 기존연구의 고찰 <ul style="list-style-type: none"> 1. 정률방법 2. 도시별 인구규모 대비 표본을 3. 카테고리수를 고려하지 않는 표본수 결정식 4. 카테고리수를 고려하지 않는 O/D 조사의 표본을 결정 III. O/D 추정을 위한 이상적인 표본을 산정방법 및 문제점 <ul style="list-style-type: none"> 1. 산정방법 | <ul style="list-style-type: none"> 2. 문제점 IV. 예산제약하에서 최소표본을 산정방법 제안 <ul style="list-style-type: none"> 1. zero cell 제외 2. 대권역으로 적용 3. 해당존이 인접존으로 통행하는 경우만 적용 4. 해당존이 다음 인접존으로 통행하는 경우까지 적용 V. 결론 참고문헌 |
|---|--|

Key Words : 기종점 통행량(O/D), 최소표본을, 예산제약, zero cell, 인접존

요 약

O/D 추정을 위한 표본조사시 최소표본을의 결정은 조사 전체 및 구축된 O/D의 신뢰성과 직결되는 중요한 문제이다. 현재 대부분의 O/D 추정을 위한 교통조사시 정해진 기준 없이 단순히 전체 모집단에 대하여 정률로 표본을을 결정하거나, 모집단의 크기에 따라 약간씩 표본수를 가감하는 표본을을 사용하고 있으나, 적용시 신뢰성 문제가 존재하므로 이에 대한 보완이 필요하다.

본 연구에서는 이러한 문제점의 해결방안으로 최악의 경우에도 zero cell을 없애도록 고안된 교통조사지침의 표본수 결정식을 이용하되, 이 방법의 문제점인 과도한 표본을을 줄이기 위하여 카테고리수를 조사여건에 따라 차등 적용하여 예산제약의 문제를 해결하는 방안에 대하여 검토하였다. 전국 지역간 여객 O/D 자료를 대상으로 기존 O/D 자료에서 zero cell을 제외하는 경우(1안), 대권역으로 적용하는 경우(2안), 인접존으로 통행하는 경우(3안), 다음 인접존까지 통행하는 경우(4안) 등 4개 안을 제안하여 분석하였고, 그 결과 각 대안들은 신뢰성과 표본을 측면에서 대체관계(trade-off)로 각각 장단점을 내포하고 있는 것으로 나타났으므로, 각 조사기관은 조사의 신뢰성과 예산 등의 문제를 포괄적으로 고려하여 최적의 방법을 선택하여 적용하여야 할 것이다.

A large sample can provide more information about the population. As the sample size increases, analysts will be more confident about the survey results. On the other hand, the costs for survey will increase in time and manpower. Therefore, determination of the sample size is a trade-off between the required accuracy and the cost. In addition, permitted error and significance level should be considered.

Sample size determination in surveys for O/D estimation is also connected with confidence of survey result. However, the past methods were usually too simple to consider confidence. Therefore, a new method for O/D surveys was proposed and it was accurate enough, but it has too large sample size when we have current budget constraint.

In this research, several minimum sample size determination methods for origin-destination survey under budget constraint were proposed. Each method decreased sample size, but has its own advantages. Selection of the sample size will depend on the study purpose and budget constraint.

1. 서론

전국 지역간 여객 기종점 통행량(O/D)은 국토종합 계획, 국가기간교통망계획을 비롯한 각종 교통계획 및 물류계획의 효과적 수립, 시행, 평가와 교통정책 개발 및 연구의 신뢰성 증진을 위한 필수적인 기초자료이다.

대규모 지역간 O/D 구축시 전수조사를 할 경우 완벽한 조사가 가능하여 신뢰성 있는 조사결과가 나올 수 있으나, 현실적인 예산제약으로 인하여 전체 모집단에 대한 조사가 불가능하므로 신뢰성을 확보할 수 있는 범위내에서 표본조사를 실시하여 표본 O/D를 구축하고, 이를 전체 모집단으로 전수화함으로써 O/D를 구축하고 있다. 따라서 O/D 표본조사시 유효표본율의 결정은 조사 전체 및 O/D의 신뢰성과 정확성에 직결되는 중요한 문제이므로 신중한 고려가 필요하다.

조사시 표본율을 증가시키면 실제 상황에 대한 묘사가 더 정확해지고 결과에 대한 확신도 증가되어 결국 신뢰성도 높아지게 된다. 다만 표본율의 증가는 시간과 인력의 증가를 초래하여 비용증가로 연결되므로 최소표본율 결정은 정확도와 비용의 trade-off 관계에 있다고 할 수 있다. 따라서 항상 예산과 신뢰성을 함께 고려해야 할 것이다.

현재 국내 여러 기관에서 정기적 또는 비정기적으로 실시되고 있는 O/D 관련 교통조사의 대부분은 일정한 기준 없이 단순히 전체 모집단에 대하여 정률로 표본율을 결정하거나, 모집단의 크기에 따라 약간씩 표본수를 가감하는 정도에 그치고 있으나, 이러한 표본율을 그대로 적용할 경우 신뢰성에 대한 문제점이 발생되므로 이에 대한 보완이 필요한 실정이다.

본 연구는 기존 표본을 결정방법을 고찰하고, 표본을 결정시 발생하는 문제점을 보완하기 위하여 현실적 예산제약이 있을 때 신뢰성을 크게 저하시키지 않는 범위내에서 O/D를 추정하기 위한 최소표본율 결정방법에 대하여 제안하고자 한다.

II. 기존연구의 고찰

먼저 유효표본율을 결정하는 방법으로 현재 사용하고 있는 정률방법, 도시별 인구규모 대비 제안된 표본수를 적용하는 방법, 카테고리수(k)를 고려하지 않은 일반적인 표본수 결정식을 적용하는 방법, 카테고리수를 고려하지 않고 O/D쌍을 적용하여 표본율을 결정하는 방법에 대해 고찰하였다.

1. 정률방법

국내 교통관련 표본조사를 하는 대부분의 경우 유효표본율을 예산에 맞추어 일률적으로 적용하는 정률방법이 이용되고 있다. 이 방법은 모집단 크기를 조사하여 주어진 예산에 맞게 표본율을 일정하게 결정하는 방법으로 지금까지 많이 이용되어 왔다. 실제 적용시 현실적이고 간편한 장점이 있으나, 통계적인 면에서 볼 때 표본수에 따른 신뢰도가 고려되지 못함으로써 조사의 신뢰성 확보에 어려움이 있는 문제점이 있다.

한국교통연구원은 「1999년 국가교통DB구축사업」의 '가구통행실태조사'시 이 방법을 적용하여 전체 가구수(모집단) 대비 7~10%에 해당하는 가구를 조사하였다. 이때 최소 유효 표본율은 약 2~5%가 되도록 표본크기를 설정하였다. 예산을 고려하여 각 도시별 조사 가구수를 산정하였으며 <표 1>과 같다.

<표 1> 각 도시별 조사가구수 산출 결과

구분	부산	대구	광주	대전	울산
총 가구수	1,157,850	833,975	399,033	407,778	307,108
조사가구수	81,050	62,548	31,923	32,622	30,711
표본율	7.00	7.50	8.00	8.00	10.00

자료: 「1999년 국가교통DB구축사업」중 '가구통행 실태조사 및 기초분석', 한국교통연구원

2. 도시별 인구규모 대비 표본율

Homburger(1992)는 도시별 인구규모에 따라 조사표본수를 적용하는 방법을 제안하였다. 이 방법은 간편하기는 하나, 가구당 인구수, 존당 표본수 등에 대해 연구에서 제안된 값을 획일적으로 사용하므로 실제 적용하고자 하는 지역의 특성과 일치하지 않을 수 있다.

서울시정개발연구원은 「2002년 가구통행실태조사」

<표 2> 도시규모별 적정표본율

도시별 인구규모(인)	존내 가구수(가구)	도시내 존의 개수 ²⁾ (개)	표본수 ²⁾ (가구)	표본율(%)
50,000	150	100	3,000	20.0
100,000	150	200	6,000	20.0
200,000	300	250	7,500	10.0
500,000	300	500	15,000	10.0
1,000,000	300	1,000	30,000	10.0
2,000,000	600	1,000	30,000	5.0
5,000,000	1,500	1,000	30,000	2.0

주: 1) 가구당 3.3인 기준

2) 존당 표본수 30개 기준

에서 이 방법을 적용하여 전체 가구수(모집단) 대비 2~20%에 해당하는 가구를 조사하였으며, 이때 최소 유효표본율은 약 2%가 되도록 표본크기를 설정하였다. 도시규모별 적정표본율은 <표 2>와 같다.

3. 카테고리수를 고려하지 않는 표본수 결정식

Rea and Parker(1997)는 기존 조사관련 통계서 적에서의 카테고리수를 고려하지 않는 일반적인 표본수 산정방법을 제안하고 있다. 즉, 신뢰구간(confidence interval), 신뢰계수(Z score for various levels of confidence), 표본오차(standard error)를 변수로 하 되, 카테고리수는 고려하지 않은 표본수 결정방식을 통하여 적절한 최소 표본크기를 제안하였다. 신뢰구간은 정확도에 따라 1~10%로 설정하며, 보통 3~5%를 사용한다. 신뢰계수는 유의수준이 99%일 경우 2.575, 95%일 경우 1.96을 적용하며, 표준오차는 식(2)에 의하여 결정된다.

모집단의 크기는 100,000을 기준으로 나뉘며, 이는 모집단의 크기가 100,000을 넘으면 모집단의 크기가 커서 복원추출에 해당된다고 가정하기 때문이다. 이때, 모집단의 크기를 도출하는 일반적인 식은 식(1)과 같으나, 모집단의 크기가 100,000이하일 경우 비복원추출에 해당한다고 가정하기 때문에 식(5)를 적용하여야 한다.

1) 모집단이 클 경우

모집단이 100,000 이상일 경우 식(1)을 적용한다.

$$C_p = \pm Z_\alpha \times \sigma_p \tag{1}$$

여기서, C_p : 신뢰구간
 Z_α : 신뢰계수
 σ_p : 표준오차

표준오차는 식(2)에 의해서 다음과 같이 산정된다.

$$\sigma_p = \sqrt{p(1-p)/n} \tag{2}$$

여기서, p : true proportion

p 는 미지의 수로, 보통 표본수가 최대가 되는 0.5를 그 값으로 한다.

$$C_p = \pm Z_\alpha \sqrt{\frac{p(1-p)}{n}} \tag{3}$$

따라서,

$$n = \left(\frac{Z_\alpha (.5)}{C_p} \right)^2 \tag{4}$$

<표 3> 큰 모집단에 대한 적절한 최소표본크기

신뢰구간 (Confidence Interval)	표본크기	
	유의수준 95%	유의수준 99%
±1	9,604	16,590
±2	4,201	4,148
±3	1,068	1,844
±4	601	1,037
±5	385	664
±6	267	461
±7	196	339
±8	151	260
±9	119	205
±10	97	116

2) 모집단이 작을 경우

모집단이 100,000 이하일 경우 식(5)를 적용한다.

$$n = \left(Z_\alpha \frac{\sqrt{p(1-p)}}{C_p} \cdot \sqrt{\frac{N-n}{N-1}} \right)^2 \tag{5}$$

따라서,

$$n = \frac{Z_\alpha^2 (.25)N}{Z_\alpha^2 (.25) + (N-1)C_p^2} \tag{6}$$

<표 4> 작은 모집단에 대한 적절한 최소표본크기

모집단(인)	샘플크기					
	유의수준 95%			유의수준 99%		
	신뢰구간			신뢰구간		
	±3%	±5%	±10%	±3%	±5%	±10%
500	250	218	81	250	250	125
1,000	500	278	88	500	399	143
1,500	624	306	91	750	460	150
2,000	696	323	92	959	489	154
3,000	788	341	94	1,142	544	158
5,000	880	357	95	1,347	586	161
1,0000	965	370	96	1,556	622	164
20,000	1,014	377	96	1,687	642	165
50,000	1,045	382	96	1,777	655	166
100,000	1,058	383	96	1,809	659	166

4. 카테고리수를 고려하지 않는 O/D 조사의 표본을 결정

Uday and Keyes(1995)와 Stokes and Chira-Chavala(1989)는 카테고리수를 고려하지 않고 특정 O-D쌍을 갖는 통행의 비율을 적용하여 표본수를 산정하는 방법을 제시하고 있다.

이 방법은 특정 O-D쌍의 비율을 가정하여 O/D 조사시 기존방법의 현실적 문제를 보완한 것으로 빈번하게 발생하는 특정 O-D쌍을 사전에 찾는 것이 어려우므로 이를 가정하게 되는데, 이에 대한 타당성을 확보하여 적용해야 하는 어려움을 수반하고 있다.

이 연구에서는 특정 O-D쌍의 비율인 p는 15% 이하, 퍼센트 error율은 ±15% 이내로 적용하고, 신뢰계수는 유의수준 90%로 가정하여 1.645를 적용하였으며, 모집단(12시간 교통량)에 따라 식(7)과 같이 적정 표본율을 산정하였다.

$$r = (Z^2 pq) / [(N-1)W^2 + (Z^2 pq)] \tag{7}$$

- 여기서, r : 표본율
- p : 특정 O-D쌍의 비율
- q : (1-p)
- W : 요구정확성(error × p)
- N : 교통량
- Z : 신뢰계수

〈표 5〉 모집단에 대한 표본율

모집단(대)	표본율(%)
5,000 이하	0.34
5,000 ~ 10,000	0.24
10,000 ~ 20,000	0.14

III. O/D 추정을 위한 이상적인 표본율 산정방법 및 문제점

1. 산정방법

O/D 추정을 위한 조사를 실시할 경우, O/D쌍의 특성상 위에서 검토된 방법들을 이용하면 사실상 zero cell이 많이 발생하게 된다. 특정 O-D쌍의 비율을 미리 알 수 없으므로 zero cell이 발생되면 전수화시 문

제가 발생하게 된다. 따라서 이를 보완하기 위한 표본율 산정방법으로 교통조사지침에서는 O-D쌍에 따른 카테고리수를 적용하여 zero cell을 없애도록 하는 표본수 결정율을 이용하고 있다. 이 식의 변수로는 모집단수, 상대오차, 신뢰계수, 카테고리수가 사용되며, 모집단수는 결정되어져 있고, 상대오차, 신뢰계수, 카테고리수는 조사비용과 조사시간 등 조사시행 여건에 부합하는 범위내에서 차등 적용이 가능하도록 하고 있다.

이 식은 zero cell이 발생하는 최악의 경우(worst case)를 방지하도록 다항분포(multinormal distribution)의 분포비율을 최소화 할 수 있는 동시상태 허용오차를 고려하여 결정되었다.

지침에 의한 표본수 결정식은 식(8)과 식(9)와 같다.

$$\text{추출율}(r) = \frac{1}{\left(\frac{\epsilon}{z}\right)^2 \times N \times \frac{1}{k-1} + 1} \tag{8}$$

$$\text{표본수}(n) = N \times r \tag{9}$$

- 여기서, ϵ : 상대오차
- z : 신뢰계수
- k : 카테고리수
- r : 추출율
- N : 모집단수

한국교통연구원은 「2001년 국가교통DB구축사업」의 「가구통행실태조사」와 「2005년 국가교통DB구축사업」의 「전국 지역간 여객 O/D조사」시 이 식을 적용하였다.

2. 문제점

이 방법은 zero cell 발생이라는 최악의 경우를 없앤다는 점에서 전수화를 위한 O/D 추정 표본율 산정방법으로 이상적이기는 하나, 최소표본율이 과대해지는 문제점이 있다. 신뢰성을 고려한 상대오차, 신뢰계수, 카테고리수를 적용할 경우 최소표본율이 과대해져 새로운 조사시 실제 적용하기에는 비현실적이다. 따라서 실제 조사에 적용하기 위해 결국 오차범위를 늘리고 신뢰도를 줄이게 되므로 결과적으로 조사 전체의 신뢰성을 저하시킬 수 있다.

실제로 최소기준이라 할 수 있는 신뢰계수는 1.67(90% 유의수준), 상대오차는 10%로 하고, 카테고리수는 전국 지역간 여객 O/D 조사의 존구분인 162개 존

(제주도·울릉도 제외)으로 적용할 경우 최소 유효표본수는 <표 6>과 같다.

<표 6> 표본수 산정식에 의한 최소 유효표본수

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	97.8	978	50,000	47.3	23,657
3,000	93.7	2,812	55,000	44.9	24,720
5,000	90.0	4,499	60,000	42.8	25,682
10,000	81.8	8,179	65,000	40.9	26,556
15,000	75.0	11,244	70,000	39.1	27,355
20,000	69.2	13,837	75,000	37.4	28,086
25,000	64.2	16,059	80,000	35.9	28,760
30,000	59.9	17,984	85,000	34.6	29,381
35,000	56.2	19,669	90,000	33.3	29,956
40,000	52.9	21,155	95,000	32.1	30,490
45,000	49.9	22,475	100,000	31.0	30,988

위의 다른 연구들에 의한 표본수와 비교해 볼 때, 표본수가 상대적으로 많아 현실적으로 적용하는데 어려움이 있음을 알 수 있다.

IV. 예산제약하에서 최소표본을 산정방법 제안

대규모 O/D 조사시 식(8)과 식(9)에 의한 표본율은 예산상의 문제로 인해 조사표본수를 확보하는 것이 현실적으로 불가능하다. 현실적인 면에서 예산을 고려하면 평균 표본율이 3~5%를 넘지 못하며, 10%를 넘기는 것은 현재의 정부예산구조상 불가능한 것으로 보인다. 따라서 신뢰성을 저하시키지 않는 범위내에서 예산 사정에 따라 최소표본율을 유동적으로 변동시키는 것이 필요한데, 이에 대한 대안으로 예산사정에 따른 보완방안을 제시하고자 한다.

예산제약을 극복하기 위한 표본산정 방법으로 교통 조사지침에 의한 표본을 산정식에서 관측에 의한 모집단수(교통량)가 결정되어 있고, 상대오차(10%) 및 신뢰계수(1.67)를 최소한의 신뢰도를 유지할 수 있게 결정할 때, 카테고리수를 기존 전국 지역간 여객 O/D 자료(2003년 기준), 권역구분, 인접존의 수 등을 이용하여 조정함으로써 표본을 산정방법에 대한 보완방안을 <표 7>과 같이 제안하였다

각 대안에 대한 비교·분석을 실시하기 위해 서울시, 천안시, 마산시 등 3개 지역을 선정하였다. 이는 대안별로 카테고리를 구분하여 고려하였을 때, 이 3개 지역이 대안별로 각기 다른 카테고리를 보이므로 특성이 뚜렷

<표 7> 각 도시별 조사가구수 산출 결과

구분	내용	카테고리수
1안	162개 존에서 통행이 없는 zero cell을 제외	존마다 다름
2안	전국을 대권역으로 나누어 해당존이 그존이 속한 권역으로 통행하는 경우만 적용	권역마다 다름
3안	해당존이 인접존으로 통행하는 경우만 적용	존마다 다름
4안	해당존이 다음 인접존으로 통행하는 경우까지 적용	존마다 다름

하고, 대표성을 가진다고 판단되기 때문이다.

1. zero cell 제외 (1안)

최소표본을 조정방법으로 과거의 O/D 자료로부터 발생한 zero cell을 제외하는 방법을 제안하였다. O/D 조사의 표본을 산정시 zero cell을 없애는 것이 이상적인 만큼 zero cell을 없애야 하나, 현실적으로 예산제약에 의해 표본율을 올리는 것이 어려우므로 과거의 O/D 자료로부터 zero cell을 찾아 이를 표본을 산정시 제외하는 방법이다. 현실적으로 존재하지 않는 O/D 쌍은 조사하지 않음으로써 표본율을 낮추는 방법이다.

한국교통연구원에서 발표한 2003년 전국 지역간 여객 O/D 자료에 의하면, 제주도와 울릉도를 제외한 본토 기준 162개 시·군에서 발생하는 zero cell의 비율은 평균 52.4%로, 각 시·군에서부터 162개 존 중 절반 이상의 존으로의 통행은 없는 것으로 나타났다. 따라서 통행이 없는 존에 대해서 카테고리수를 제외시킴으로써 표본수를 줄일 수 있었다. 각 존에 따라 zero cell의 비율이 다르므로 적용되는 카테고리수 역시 달라진다.

서울시의 경우 다른 지역에 비해 zero cell의 비율

<표 8> k=96일 경우 추출율 및 표본수(서울시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	96.4	964	50,000	34.6	17,318
3,000	89.8	2,695	55,000	32.5	17,881
5,000	84.1	4,206	60,000	30.6	18,379
10,000	72.6	7,260	65,000	29.0	18,822
15,000	63.9	9,578	70,000	27.5	19,220
20,000	57.0	11,397	75,000	26.1	19,578
25,000	51.5	12,863	80,000	24.9	19,903
30,000	46.9	14,069	85,000	23.8	20,199
35,000	43.1	15,080	90,000	22.7	20,469
40,000	39.8	15,938	95,000	21.8	20,717
45,000	37.1	16,676	100,000	20.9	20,945

이 적고, 천안시는 보통 수준이며, 마산시는 zero cell의 비율이 큰 지역으로 나타났으므로 이 세 지역을 비교대상으로 선정하여 비교·분석하였다.

예를 들어 서울시는 zero cell이 66개로 나머지 95개 존으로만 이동하였으므로 카테고리 수는 96개(서울시 포함)로 적용하였다.

zero cell을 제외시킨 결과, 표본 수는 모집단에 따라 약 1%~32% 감소하였다.

천안시는 zero cell이 75개로 나머지 86개 존으로만 이동하였으므로 카테고리 수는 87개(천안시 포함)로 적용하였다.

〈표 9〉 k=87일 경우 추출율 및 표본수 (천안시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	96.0	960	50,000	32.4	16,209
3,000	88.9	2,666	55,000	30.4	16,701
5,000	82.7	4,137	60,000	28.6	17,135
10,000	70.6	7,057	65,000	27.0	17,520
15,000	61.5	9,228	70,000	25.5	17,864
20,000	54.5	10,906	75,000	24.2	18,173
25,000	49.0	12,241	80,000	23.1	18,452
30,000	44.4	13,329	85,000	22.0	18,706
35,000	40.7	14,232	90,000	21.0	18,938
40,000	37.5	14,994	95,000	20.2	19,150
45,000	34.8	15,646	100,000	19.3	19,345

zero cell을 제외시킨 결과, 표본수는 모집단에 따라 약 1%~38% 감소하였다.

마산시는 zero cell이 113개로 거의 70% 존으로의 이동은 없었다. 따라서 카테고리수는 49개(마산시 포함)로 다른 지역에 비해 적었다.

zero cell을 제외시킨 결과, 표본 수는 모집단에 따

〈표 10〉 k=49일 경우 추출율 및 표본수 (마산시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	93.0	930	50,000	21.1	10,560
3,000	81.7	2,451	55,000	19.6	10,766
5,000	72.8	3,640	60,000	18.2	10,945
10,000	57.2	5,724	65,000	17.1	11,101
15,000	47.2	7,074	70,000	16.1	11,238
20,000	40.1	8,019	75,000	15.1	11,359
25,000	34.9	8,718	80,000	14.3	11,468
30,000	30.9	9,256	85,000	13.6	11,565
35,000	27.7	9,683	90,000	12.9	11,653
40,000	25.1	10,030	95,000	12.4	11,733
45,000	22.9	10,317	100,000	11.8	11,806

라 약 3%~43% 감소하였다.

결과적으로 카테고리수가 적어질수록 표본율도 줄었으며, 모집단이 클수록 표본율의 감소폭이 커지는 것으로 나타났다.

이 방법을 적용할 경우 표본율은 기존 경우보다 줄어 드나, 표본율이 아직 높아 현실적인 예산상의 문제는 여전히 나타나게 되므로 예산상의 제약이 적을 경우 고려할 수 있는 대안이다. 또한 향후 zero cell에 어떤 요인에 의해 통행이 유발될 수도 있다는 점에 유의하여야 한다.

2. 대권역으로 적용 (2안)

예산제약을 조금 더 반영하기 위하여 전국을 수도권, 충청권, 강원권, 경상권, 전라권의 5개 권역으로 나누어 해당존이 그 존이 속한 권역으로 통행하는 경우가 대부분이라고 가정(zero cell에 가까운 존은 zero cell로 가정)하여 그 경우만 카테고리수로 반영하는 방법을 적용하였다. 각 권역내에 속하는 중존수(162개 기준)가 카테고리수가 되며, 그 수는 〈표 11〉과 같다.

〈표 11〉 권역별 카테고리수

권역	카테고리수(개)
수도권(서울, 인천, 경기도)	33
충청권(대전, 충청남·북도)	29
강원권(강원도)	18
경상권(대구, 부산, 울산, 경상남·북도)	45
전라권(광주, 전라남·북도)	37

서울시는 수도권역, 천안시는 충청권역, 마산경상권역에 속하므로 이 세 지역에 대해 해당 권역의 카테고리수를 적용하여 분석하였다.

서울시의 카테고리수는 수도권의 경우인 33개가 되

〈표 12〉 k=33일 경우 추출율 및 표본수 (서울시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	89.9	899	50,000	15.1	7,573
3,000	74.8	2,245	55,000	14.0	7,679
5,000	64.1	3,205	60,000	12.9	7,769
10,000	47.2	4,716	65,000	12.1	7,847
15,000	37.3	5,595	70,000	11.3	7,915
20,000	30.9	6,171	75,000	10.6	7,975
25,000	26.3	6,577	80,000	10.0	8,029
30,000	22.9	6,878	85,000	9.5	8,076
35,000	20.3	7,111	90,000	9.0	8,119
40,000	18.2	7,297	95,000	8.6	8,158
45,000	16.5	7,447	100,000	8.2	8,193

며, 이를 적용한 모집단별 표본수는 <표 12>와 같다.

서울시의 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 8%~73% 감소하였고, 1안에 비해서 많이 감소하였다고 볼 수 있다.

천안시의 카테고리수는 충청권의 경우인 29개이며, 이를 적용한 모집단별 표본수는 <표 13>과 같다.

<표 13> k=29일 경우 추출 및 표본수 (천안시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	88.6	886	50,000	13.5	6,754
3,000	72.2	2,167	55,000	12.4	6,838
5,000	61.0	3,048	60,000	11.5	6,910
10,000	43.8	4,385	65,000	10.7	6,971
15,000	34.2	5,135	70,000	10.0	7,025
20,000	28.1	5,616	75,000	9.4	7,073
25,000	23.8	5,950	80,000	8.9	7,114
30,000	20.7	6,196	85,000	8.4	7,152
35,000	18.2	6,384	90,000	8.0	7,185
40,000	16.3	6,533	95,000	7.6	7,216
45,000	14.8	6,654	100,000	7.2	7,243

천안시의 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 9%~76% 감소하였다.

마산시의 카테고리수는 경상권의 경우인 45개이며, 이를 적용한 모집단별 표본수는 <표 14>와 같다.

<표 14> k=45일 경우 추출 및 표본수 (마산시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	92.5	925	50,000	19.7	9,853
3,000	80.4	2,411	55,000	18.2	10,033
5,000	71.1	3,553	60,000	17.0	10,188
10,000	55.1	5,510	65,000	15.9	10,322
15,000	45.0	6,750	70,000	14.9	10,441
20,000	38.0	7,605	75,000	14.1	10,546
25,000	32.9	8,231	80,000	13.3	10,639
30,000	29.0	8,709	85,000	12.6	10,723
35,000	26.0	9,086	90,000	12.0	10,799
40,000	23.5	9,390	95,000	11.4	10,867
45,000	21.4	9,642	100,000	10.9	10,930

마산시의 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 5%~64% 감소하였다.

이 방법은 기존의 문제점인 예산제약의 문제를 많이 해소하였으나, 같은 권역일 경우 무조건 같은 카테고리수를 적용하므로 권역내 각 지역의 특성이 고려되지 않을 수 있다는 문제점이 있다.

3. 해당존이 인접존으로 통행하는 경우만 적용 (3안)

예산제약이 심한 경우 카테고리수를 줄이는 것이 필수불가결하다. 이 경우 각 존은 여러 존들로 둘러싸여 있으므로 인접한 존으로 통행하는 경우가 대부분이라고 가정하여 인접존의 수를 카테고리수로 적용하는 방법이다. 각 존에 따라 인접존 수가 다르므로 적용되는 카테고리수 역시 달라진다.

서울시의 경우 타 지역에 비해 인접 존의 수가 많고, 천안시는 보통 수준이며, 마산시는 적은 것으로 나타났다. 이 세 지역을 비교대상으로 선정하여 비교·분석하였다.

서울시는 인접 존이 13개, 천안시는 7개, 마산시는 4개이며, 각 카테고리별 추출 및 표본수는 <표 15>~<표 17>에 나타나 있다.

<표 15> k=13일 경우 추출 및 표본수 (서울시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	77.0	770	50,000	6.3	3,137
3,000	52.7	1,582	55,000	5.7	3,155
5,000	40.1	2,005	60,000	5.3	3,170
10,000	25.1	2,507	65,000	4.9	3,183
15,000	18.2	2,736	70,000	4.6	3,194
20,000	14.3	2,867	75,000	4.3	3,204
25,000	11.8	2,952	80,000	4.0	3,212
30,000	10.0	3,011	85,000	3.8	3,220
35,000	8.7	3,055	90,000	3.6	3,227
40,000	7.7	3,088	95,000	3.4	3,233
45,000	6.9	3,115	100,000	3.2	3,238

서울시 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 21%~89% 감소하였다.

<표 16> k=7일 경우 추출 및 표본수 (천안시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	62.6	626	50,000	3.2	1,619
3,000	35.8	1,074	55,000	3.0	1,624
5,000	25.1	1,254	60,000	2.7	1,628
10,000	14.3	1,433	65,000	2.5	1,631
15,000	10.0	1,505	70,000	2.3	1,634
20,000	7.7	1,544	75,000	2.2	1,637
25,000	6.3	1,568	80,000	2.0	1,639
30,000	5.3	1,585	85,000	1.9	1,641
35,000	4.6	1,597	90,000	1.8	1,643
40,000	4.0	1,606	95,000	1.7	1,644
45,000	3.6	1,613	100,000	1.6	1,646

천안시 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 36%~94% 감소하였다.

〈표 17〉 k=4일 경우 추출율 및 표본수 (마산시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	45.6	456	50,000	1.6	823
3,000	21.8	654	55,000	1.5	824
5,000	14.3	717	60,000	1.4	825
10,000	7.7	772	65,000	1.3	826
15,000	5.3	792	70,000	1.2	827
20,000	4.0	803	75,000	1.1	827
25,000	3.2	810	80,000	1.0	828
30,000	2.7	814	85,000	1.0	829
35,000	2.3	817	90,000	0.9	829
40,000	2.0	820	95,000	0.9	829
45,000	1.8	821	100,000	0.8	830

마산시 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 53%~97% 감소하였다.

이 방법은 예산제약이 심할 때 예산제약의 문제를 거의 해결할 수 있으나, 존이 클수록 둘러싸는 존이 많을 확률이 크므로 존이 클수록 통행이 많은 것으로 간주하게 된다. 따라서 존 크기는 크지만 통행이 별로 없는 존의 경우 표본수가 과대 산정될 수 있다. 또한 부산시와 같이 바다와 접한 경우 카테고리수가 적어지게 되는 문제가 있다. 현실적으로 예산제약이 심하지 않다면 신뢰성 차원에서 권장하지 않는 방법이다.

4. 해당존이 다음 인접존으로 통행하는 경우까지 적용(4안)

3안의 경우 표본율은 작아지나 실제 통행의 대부분이 인접 존으로 발생한다는 가정은 비현실적인 면이 있다. 4안은 3안과 비슷한 방법으로 인접한 존의 인접한 존까지 한 단계 더 적용하여 zero cell에 가까운 존의 수를 줄이는 방법이다.

서울시는 해당 존이 25개, 천안시는 22개, 마산시는 13개이며, 각 카테고리별 추출율 및 표본수는 〈표 18〉~〈표 20〉에 나타나 있다.

서울시 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 11%~79% 감소하였으며, 3안과 비교하여 약 13%~97% 증가하였다.

〈표 18〉 k=25일 경우 추출율 및 표본수 (서울시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	87.0	870	50,000	11.8	5,903
3,000	69.1	2,072	55,000	10.8	5,967
5,000	57.2	2,862	60,000	10.0	6,022
10,000	40.1	4,010	65,000	9.3	6,068
15,000	30.9	4,628	70,000	8.7	6,109
20,000	25.1	5,015	75,000	8.2	6,145
25,000	21.1	5,280	80,000	7.7	6,177
30,000	18.2	5,472	85,000	7.3	6,205
35,000	16.1	5,619	90,000	6.9	6,230
40,000	14.3	5,734	95,000	6.6	6,253
45,000	12.9	5,827	100,000	6.3	6,273

천안시 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 3%~82% 감소하였으며, 3안과 비교하여 약 36%~236% 증가하였다.

〈표 19〉 k=22일 경우 추출율 및 표본수 (천안시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	85.4	854	50,000	10.5	5,243
3,000	66.1	1,984	55,000	9.6	5,293
5,000	53.9	2,697	60,000	8.9	5,336
10,000	36.9	3,694	65,000	8.3	5,373
15,000	28.1	4,212	70,000	7.7	5,405
20,000	22.7	4,530	75,000	7.2	5,432
25,000	19.0	4,745	80,000	6.8	5,457
30,000	16.3	4,900	85,000	6.4	5,479
35,000	14.3	5,017	90,000	6.1	5,499
40,000	12.8	5,109	95,000	5.8	5,517
45,000	11.5	5,182	100,000	5.5	5,533

마산시 표본수는 k=162일 경우와 비교하여 모집단에 따라 약 21%~89% 감소하였으며, 3안과 비교하

〈표 20〉 k=13일 경우 추출율 및 표본수 (마산시)

모집단 (대)	추출율 (%)	표본수 (개)	모집단 (대)	추출율 (%)	표본수 (개)
1,000	77.0	770	50,000	6.3	3,137
3,000	52.7	1,582	55,000	5.7	3,155
5,000	40.1	2,005	60,000	5.3	3,170
10,000	25.1	2,507	65,000	4.9	3,183
15,000	18.2	2,736	70,000	4.6	3,194
20,000	14.3	2,867	75,000	4.3	3,204
25,000	11.8	2,952	80,000	4.0	3,212
30,000	10.0	3,011	85,000	3.8	3,220
35,000	8.7	3,055	90,000	3.6	3,227
40,000	7.7	3,088	95,000	3.4	3,233
45,000	6.9	3,115	100,000	3.2	3,233

여 약 69%~290% 증가하였다.

이 방법은 3안과 같은 문제점을 내포하고 있으나, 통행지역의 범위를 늘림으로써 조사의 신뢰성을 높일 수 있는 장점이 있다. 현실적인 면에서 볼 때 현재의 예산구조하에서 가장 현실성이 있다고 보여진다.

V. 결론

현재 여러 기관에서 정기적 또는 비정기적으로 실시되고 있는 교통조사의 대부분은 정해진 기준 없이 단순히 전체 모집단에 대하여 정률로 유효표본율을 결정하거나, 모집단 크기에 따라 약간씩 표본수를 가감하는 표본율을 사용하고 있다. 이러한 표본율 산정방법은 O/D 추정을 위한 대규모 조사에 그대로 적용할 경우 신뢰성에 큰 문제가 발생할 수 있다. 반면 현재 교통조사지침에서 제안하는 방법은 신뢰성을 높이는 방법으로는 이상적이나 표본율이 너무 높아 현실적으로 이용하기에 제약이 있으므로 이에 대한 보완이 필요하다.

교통관련 표본조사시 표본율의 결정방법은 앞서 언급한 것처럼 조사전체 및 O/D의 신뢰성에 큰 영향을 주나 발전된 방법으로 제안된 조사지침서는 대부분 통계적으로 유의한 반면 현실성이 떨어지는 문제가 있었다. 실제 대규모 조사에서는 사실상 예산제약이 항상 존재하므로 보다 현실적인 표본율 결정방법이 필요하다. 실제로 국내에서 가장 대규모 O/D 조사를 실시하는 한국교통연구원의 과거 조사 표본율을 살펴본 결과, 예산의 제약으로 인하여 낮은 정률의 표본율로 예산에 맞추어 실시한 것으로 나타났으며, 앞으로도 현재와 같은 예산규모로는 조사지침의 표본율 산정방법을 이용할 경우 신뢰성을 확보할 수 있는 표본율을 구하는 것이 어려운 것으로 분석되었다.

본 연구에서는 이러한 문제점에 대한 해결방안으로 교통조사지침에서 제시한 표본수 결정식을 이용하되, 카테고리수를 조사여건에 따라 차등 적용하여 신뢰성을 낮추지 않으면서 표본율은 낮추어 예산제약의 문제를 해결하는 4개 방안에 대하여 검토해 보았다.

카테고리수의 산정시 167개존 중에 O/D를 이용하여 zero cell을 제거하여 적용한 결과(1안), 통행패턴을 반영할 수 있어 합리적이었으나, 현실적인 예산문제를 계속 수반하였고, 전국을 대권역으로 묶어 적용한 결과(2안), 현실적인 예산문제를 많이 해소하였으나 지역특성을 무시될 수 있어 문제점으로 지적되었다. 또한 인접 존으로의 통행만 적용한 결과(3안), 예산문제는 해소되었으나, 존 크기와 통행량에 대한 관계 규명이 미흡하였으며, 다음 인접 존

까지의 통행을 적용한 결과(4안), 3안 보다 신뢰성을 높일 수 있다는 점에서 가장 나은 방법이라고 할 수 있으나 3안과 동일한 문제점을 내포하는 한계점이 존재한다.

최적의 대안은 예산을 증대하여 표본율을 올리는 것이겠으나 현실적으로 어려우므로 본 연구에서는 몇가지 대안들을 제시하였다. 위의 대안들은 각각 단점과 장점을 함께 내포하고 있다. 따라서 각 조사기관은 조사의 신뢰성과 예산의 문제를 포괄적으로 고려하여 여러 가지 방안 중 최적의 방법을 선택하여 적용하는 것이 필요하다.

참고문헌

1. 한국교통연구원(2000), 「1999년 국가교통DB구축사업」 중 '가구통행실태조사 및 기초분석'.
2. 한국교통연구원(2002), 「2001년 국가교통DB구축사업」 중 '개인통행실태조사'.
3. 한국교통연구원(2005), 「2005년 국가교통DB구축사업」, 중간보고서.
4. 서울시정개발연구원(2003), '2002 서울시 가구통행실태조사'.
5. 건설교통부(2003), 교통조사지침.
6. Homburger, W. S.(1992), Fundamentals of Traffic Engineering, Institute of Transportation Studies, University of California at Berkeley.
7. Rea, L. M. and R. A. Parker(1997), Designing and Conducting Survey Research, Jossey-Bass.
8. Virkud, U. and C. S. Keyes(1995), Design and Implementation of a Statewide Roadside Origin-Destination Survey in Vermont, Transportation Research Record 1447.
9. Stokes, R. W. and T. Chira-Chavala(1989), Design and Implementation of Intercity Origin-Destination Surveys, Paper Prepared for 68th Annual Meeting of Transportation Research Board.

✉ 주 작 성 자 : 신희철

✉ 교 신 저 자 : 신희철

✉ 논문투고일 : 2006. 2. 25

✉ 논문심사일 : 2006. 4. 21 (1차)

2006. 4. 24 (2차)

✉ 심사판정일 : 2006. 4. 24

✉ 반론접수기한 : 2006. 9. 30