

A Review on Three Cards Method

Kyoung-Ho Choi¹⁾

Abstract

In this paper, we review three cards method by Droitcour et al.(2001). And we compare indirect survey methods - randomized response technique, item count method, nominative technique.

Keywords : indirect survey method, randomized response technique, sensitive issues, three cards method

1. 서론

사회조사 시 민감사안(sensitive issues)에 대하여 직접질문(direct question)을 하게 되면, 응답자는 신분노출 등이 우려되어 정직한 응답을 회피하게 된다. 이는 비표본오차(non-sampling errors)의 증대를 가져와, 많은 경우 과소추정(under estimation)이 이루어지는 요인이 된다.

그래서 이러한 문제를 해결하기 위한 많은 선행된 노력이 있어온 바, Warner(1965)의 확률화응답기법(randomized response technique : RRT), Droitcour et al.(1991)의 item count(IC)방법, Sirken(1975)과 Sudman et al.(1977)의 nominative technique(NT) 등이 이에 대한 예에 해당된다. 이러한 기법들은 조사과정에서 응답자의 익명성이 보장되어 신분보호(privacy protection)가 이루어지도록 구성되어 있다.

그러나 RRT의 경우 확률장치를 통한 응답수행과정에 대해서 응답자가 이를 이해하는데 어려움이 있으며, 나아가 여전히 정직한 응답에 따른 신분노출에 대하여 의심이나 불안감이 남아 있게 된다. 한편 NT의 경우에 있어서도 응답자가 갖는 주위사람에 대한 정보의 한계로 인하여, 정확한 추정이 이루어진다고 보장하기 어렵다.

응답자의 신분보호를 통하여 민감사안에 대한 정확한 추정을 수행하는 것은, 특정분야에 대한 국가의 정책수립에 요구되기도 한다.

우리 나라의 경우 1990년대 이후 내국인의 3D현상이 심화되면서 산업연수생이라는 명목으로 많은 외국인 노동자가 국내로 유입되어 생산현장에 투입되었다. 그런데 이들 중 상당수가 법적 테두리에서 벗어나 불법체류자 상태에 있고, 또한 이에 편승한 밀입국자가 증가되어 많은 사회문제를 야기하고 있는 실정이다. 법무부 자료에 의하면 2000년 8월 현재 국내에 거주하는 불법체류자 수는 17만 2501명으로 집계되고 있으며, 경찰청 통계에 따르면 2002년 4월 현재 불법체류 외국인 수는 26만 8천 258명으로 2001년 4월의 20만 5천 817명에 비하여 1년만에 30.3%나 급증하고 있는

1) Associate Professor, Major of Statistics, Jeonju University, Chonbuk, 560-759, Korea.
E-mail : ckh414@jeonju.ac.kr

것으로 나타났다. 이러한 불법체류자의 증가는 각종 범죄나 인권유린 등의 많은 사회문제를 야기하는 바 이에 대응하는 적절한 정책을 수립해야 하는데, 그러기 위해서는 이들에 대한 정확한 숫자를 파악하는 것이 무엇보다도 중요하다. 그런데 각 해당부처에서 발표하는 통계를 보면 서로 상이하여 정확한 숫자를 파악하기가 어렵다. 그 이유는 외국인을 대상으로 한 '불법체류' 여부에 관한 조사가 민감사안에 해당하므로, 정확한 조사가 수행되기가 어려우며 나아가 과소추정되는 데 기인한 것으로 여겨진다.

이에 본 논문에서는 이와 같은 민감사안에 대한 조사 시, 기존의 방법에 비하여 응답자의 신분보호의 정도가 높으며 응답자가 조사과정을 이해하기 쉽고 추가적인 조사가 가능한 방법으로 Droitcour et al.(2001)에 의하여 제안된 three card(TC)방법을 소개하고자 한다. 또한 아직까지는 국내에 잘 알려지지 않은 이들 간접조사기법들의 장단점을 비교하여 조사목적에 맞는 적절한 방법을 선택하는데 도움을 주고자 한다.

2. Three Card 방법

2.1 카드의 구성과 응답획득과정

TC방법은 민감사안에 대한 조사 시, 응답자의 신분보호를 통하여 신뢰성 있는 응답을 획득하기 위하여 고안된 간접조사기법이다. 이를 수행하기 위해서 먼저 크기가 각각 $n_i(i=1,2,3, \sum n_i = n)$ 인 3개의 독립표본을 추출한다. 이 때 각 표본에 속하는 응답자는 중복되지 않도록 해야한다. 다음 3개의 박스(box)로 구성된 외형과 크기가 같은 3장의 카드를 준비한다. 박스구성 시, 민감질문은 민감하지 않은 혹은 덜 민감한 질문과 같은 박스에 위치하도록 하며, 3장의 카드에 나타나는 민감질문의 내용은 같게 한다. 다만 각 카드에 대하여 각 박스에 위치되는 질문의 순서는 다르게 한다. 이제 준비가 되었으면 하나의 표본에 대하여 임의의 한 카드만을 이용하여 내용이 서로 배반인 질문들로 구성된 A, B, C 세 개의 박스 중 응답자 본인에 해당하는 박스만을 응답토록 한다. 이 때 민감질문은 민감하지 않은 질문과 함께 박스 B에 위치토록 하며, 응답자는 자신에게 해당하는 박스만을 응답토록 함으로써 응답자의 신분보호가 이루어지게 된다. 모든 응답이 이루어지게 되면 각 표본에서 민감하지 않은 집단(각 카드의 박스 A에 해당하는 집단)의 비율이 추정되며, 전체에서 이들 추정값을 뺀으로써 민감사안에 해당하는 집단에 대한 비율을 간접 추정할 수 있게 된다. 즉 TC방법을 이용한 추정은 민감하지 않은 집단에 대해서는 직접추정이 이루어지고, 민감 집단에 대해서는 간접추정이 이루어진다.

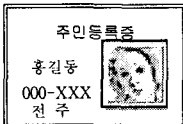



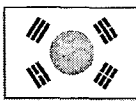







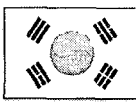
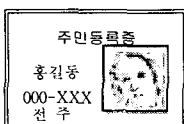
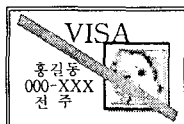



한편 TC방법은 수행과정에서 응답자의 이해를 돕고 응답선택에 도움을 주기 위하여 적절한 그림을 보조로 이용하게 된다.

국내에 거주하는 불법체류자의 비율을 추정하기 위한 다음의 예를 보도록 하자. 먼저 국내에 거주하는 외국인에 대하여, 구성요소가 다른 각각 크기 $n_i(i=1,2,3, \sum n_i = n)$ 인 3개의 독립표본을 추출한다. 다음으로 그림 2.1과 같이 구성된 3장의 카드를 준비한다.

카드 1

카드 2

카드 3

<p>A</p> <p>국적취득 (귀화, 국적이주 노동자)</p> 	<p>A</p> <p>장기체류 (상사주제원, 국제결혼, 국적미취득의 이민노동자)</p> 	<p>A</p> <p>일시체류 (학생, 방문, 관광등의 단기방문자)</p>  <p>망명, 난민 (국제법에 따른정치적 망명이나 난민)</p> 
<p>B</p> <p>장기체류 (상사주제원, 국제결혼, 국적미취득의 이민노동자)</p>  <p>일시체류 (학생, 방문, 관광등의 단기방문자)</p>  <p>불법체류 (방문기간 경과, 방문목적외 취업등)</p>  <p>망명, 난민 (국제법에 따른정치적 망명이나 난민)</p> 	<p>B</p> <p>국적취득 (귀화, 국적이주 노동자)</p>  <p>일시체류 (학생, 방문, 관광등의 단기방문자)</p>  <p>불법체류 (방문기간 경과, 방문목적외 취업등)</p>  <p>망명, 난민 (국제법에 따른정치적 망명이나 난민)</p> 	<p>B</p> <p>장기체류 (상사주제원, 국제결혼, 국적미취득의 이민노동자)</p>  <p>국적취득 (귀화, 국적이주 노동자)</p>  <p>불법체류 (방문기간 경과, 방문목적외 취업등)</p> 
<p>C</p> <p>기타 (위의 A,B 범주에 속하지 않는경우)</p> 	<p>C</p> <p>기타 (위의 A,B 범주에 속하지 않는경우)</p> 	<p>C</p> <p>기타 (위의 A,B 범주에 속하지 않는경우)</p> 

[그림 2.1] 구성된 3장의 카드

가상적인 예로 카드 1에 응답한 표본 1중 39%의 응답자가 박스 A(귀화)라고 응답했고, 카드 2에 응답한 표본 2중 20%가 박스 A(영주이민)라고 응답했으며, 카드 3에 응답한 표본 3중 4%가 박스 A(일시체류)라고 응답했다고 하자. 한편 박스 C라고 응답한 응답자가 1%라면, 우리가 알고자 하는 불법체류자의 비율은 $100\% - (39\% + 20\% + 4\% + 1\%) = 36\%$ 로 추정할 수 있다. 이 때 각 카드의 박스 A에 해당하는 집단은 직접 추정되며, 불법체류자의 비율만이 응답자의 신분노출 없이 간접 추정된다.

2.2 추정량의 분산과 표본배분

TC방법에 의하여 간접 추정되는 민감집단(불법체류집단)의 비율(P_x)의 추정량에 대한 분산식을 유도하기 위하여 먼저 다음을 정의하자.

P_1 : 국내 외국인 중 귀화한 비율

P_2 : 국내 외국인 중 영주이민자 비율

P_3 : 국내 외국인 중 일시체류자 비율

이제 $p_i(i=1,2,3)$ 를 각각 P_i 의 추정량이라 하고, p_x 를 P_x 의 추정량이라 하자. 한편 각 표본의 크기를 $n_i(i=1,2,3, \sum n_i = n)$ 라 하면, P_x 의 간접추정량과 이의 분산은 각각 다음과 같다.

$$\begin{aligned} p_x &= 1 - (p_1 + p_2 + p_3) \\ V(p_x) &= V(p_1) + V(p_2) + V(p_3) \\ &= \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2} + \frac{P_3 Q_3}{n_3} \end{aligned} \quad (2.1)$$

한편, $n_1 = n_2 = n_3 = n/3$ 이고, $P_x = P_1 = P_2 = P_3$ 라면 $V(p_x) = \frac{9(P_x Q_x)}{n}$ 이다.

위에서 알 수 있듯이 TC방법은 신뢰성 있는 응답을 얻을 수 있는 반면, 다른 간접기법에서와 같이 추정량의 분산이 증가하는 단점이 있다. 그러나 사전지식에 기초하여 P_i 중 분산이 클 것으로 예상되는 표본군의 크기는 크게 하고, 분산이 작을 것으로 예상되는 표본군의 크기는 작게 함으로써 전체적으로 식 2.1로부터 구해지는 민감집단에 대한 추정량의 분산을 줄일 수 있다. 나아가 층화표본 등을 이용함으로써 추정량의 분산을 줄일 수도 있다.

3. 간접 조사방법들의 비교

사회조사 시 개인의 사적인 문제와 관련된 민감한 사안에 대한 조사는 응답오차 등의 비표본오차와 관련된 부분으로, 오랫동안 연구의 대상이 되어왔다. 그래서 응답자의 신분보호를 이루어 신뢰성 있는 응답을 획득할 수 있는 다양한 방법들이 개발되어 온 바, RRT, IC방법, NT, TC방법 등이 이에 해당된다.

먼저, RRT는 이미 국내에도 잘 알려진 바 추가적인 설명은 생략하도록 하겠다. Sirken(1975)과 Sudman(1977)의 nominative technique(NT)방법은 대동소이하므로, Sudman et al.(1977)을 중심으로 개략적으로 살펴보면 다음과 같다. 조사하고자 하는 민감사안이 '마리화나 흡연자 비율'이라면, 추출된 표본에 대하여 "당신의 가장 친한 3명의 친구 중 마리화나를 피워본 경험이 있는 사람은 몇 명인가? 와 같은 식으로 질문을 하여 민감집단의 비율을 추정하는 방식이다. 이렇게 함으로써 응답자 자신의 마리화나 흡연여부를 묻지 않게 되어 응답자의 신분보호를 이루면서 민감집단의 비율을 추정할 수 있게 된다. Droitcour et al.(1991)의 item count(IC)방법의 기본절차를 살펴보면 다음과 같다. 먼저 두 개의 부차표본(subsample)을 추출하여 부차표본 A에 대해서는 민감항목이 포함된 4개의 항목으로 구성된 리스트를 보여주고, 부차표본 B에 대해서는 민감항목이 제거된 3

개의 항목으로 구성된 리스트를 보여주어 각각 수행한 수를 응답케 한 후, 두 부차표본의 응답을 비교함으로써 민감항목에 해당하는 집단의 비율을 추정하는 방법이다. 이중 리스트(double lists)에 의한 IC방법에 대한 추정량과 분산은 다음과 같다.

$$\hat{p} = \frac{1}{2} [(\bar{X}_{4A} - \bar{Y}_{3A}) + (\bar{Y}_{4B} - \bar{X}_{3B})]$$

$$V(\hat{p}) = \frac{1}{4} [V(\bar{X}_{4A}) + V(\bar{X}_{3A}) - 2r_{X_4 \cdot Y_3} \sqrt{V(\bar{X}_{4A})} \sqrt{V(\bar{X}_{3A})} + V(\bar{X}_{4B}) + V(\bar{X}_{3B}) - 2r_{Y_4 \cdot X_3} \sqrt{V(\bar{X}_{4B})} \sqrt{V(\bar{X}_{3B})}] \quad (3.1)$$

단 여기서 \bar{X}_{4A} (\bar{X}_{3A})는 4(3)개의 항목으로 구성된 부차표본 A에 대하여 수행했다고 응답한 수의 평균이며, \bar{X}_{4B} (\bar{X}_{3B})는 4(3)개의 항목으로 구성된 부차표본 B에 대하여 수행했다고 응답한 수의 평균이다. 나아가 $r_{X_4 \cdot Y_3}$ 는 4개의 항목으로 이루어진 부차표본 A에 대한 응답과 3개의 항목으로 이루어진 부차표본 B에 대한 응답간의 상관계수이다. 그리고 $r_{Y_4 \cdot X_3}$ 는 4개의 항목으로 이루어진 부차표본 B에 대한 응답과 3개의 항목으로 이루어진 부차표본 A에 대한 응답간의 상관계수이다

간접조사기법들이 공통적으로 갖는 장점은 두말할 나위 없이 응답자의 신분보호이며, 단점은 추정량의 분산이 직접질문에 비하여 증가한다는 점이다. 이제 이러한 기본적인 사항 이외의 이들 간접조사기법들의 대표적인 장단점을 비교해 보도록 하자.

< 표 3.1 > 간접조사기법들의 장단점 비교

	장 점	단 점
확률화 응답기법 (RRT)	<ul style="list-style-type: none"> · 응답자에 대한 사전지식이 별로 요구되지 않는다 · 전화조사에도 적용이 가능. · 질적조사뿐만 아니라 양적조사도 가능 	<ul style="list-style-type: none"> · 민감집단에 속하는 응답자의 응답부담은 줄여주는 반면, 비민감집단에 속하는 응답자의 응답부담이 증가될 수 있다 · 확률장치 제작, 조사과정을 응답자에게 이해시키기 등 시행상의 복잡성
Nominative technique (NT)	<ul style="list-style-type: none"> · 시행상의 간결성 · 추정량의 분산을 줄일 수 있다 	<ul style="list-style-type: none"> · 응답자의 친구에 대한 정보가 불확실한 경우 신뢰성 낮은 응답이 획득되어 비표본오차의 증대를 유발
Item count method (IC)	<ul style="list-style-type: none"> · 응답획득 과정이 간결 · 응답자 느끼는 혼동의 여지가 적다 · NT와 달리 응답자 자신에 대한 정보만 있으면 된다 	<ul style="list-style-type: none"> · 일부의 부차표본에 대해서만 민감항목이 질문되므로 추정량의 분산이 매우 크다
Three card method (TC)	<ul style="list-style-type: none"> · 박스 A를 응답한 응답자에 대해서는 추가적인 질문이 가능 · 효율적인 설계와 표본배분으로 추정량의 분산을 줄일 수 있다 	<ul style="list-style-type: none"> · 전화조사에 부적합 · 서로 중복되지 않는 3그룹의 표본을 추출해야 한다 · 서로 배반이고 포괄적인 집단의 구성

4. 결론 및 제언

본 논문에서는 응답자의 신분보호를 이루어 신뢰성 있는 응답을 획득할 수 있도록 개발되어 온 다양한 방법들 예컨대 - RRT, IC방법, NT, TC방법 등에 대해서 고찰해 보았다. 특히 TC방법을 주로 소개하였는데 이 방법의 적용을 통한 조사수행 가능 예로, 첫째 낙태나 약물남용 등과 같은 개인적인 문제, 둘째 폐기물 불법매립이나 탈세 등과 같은 범법행위, 셋째 운전 중에 일어나는 위반문제, 넷째 공무원의 지시 불이행의 문제 등을 들 수 있다.

실제 많은 사회조사에서 활용되고 있는 이와 같은 간접조사기법들은 나름대로 장단점이 있는 바, 주어진 여건에 맞는 방법을 선택하여 조사를 수행하는 것이 바람직하겠다.

참고문헌

- [1] Droitcour, J., Caspar, R. A., Hubbard, M. L. and Ezzati, T. M.(1991), *The item count technique as a method of indirect questioning : A review of its development and a case study application in Paul Biemer et al.*, Measurement Error in Surveys, Wiley.
- [2] Droitcour, J., Larson, E. M. and Scheuren, F. J.(2001), The three card method : Estimating sensitive survey items-with permanent anonymity of response, *Proceedings of the Annual Meeting of American Statistical Association*, August 5-9.
- [3] Sirken, M.(1975), *Network surveys of rare and sensitive conditions*, *Advances in Health Survey Research Methods*, Hyattsville, Maryland : National Center for Health Statistics.
- [4] Sudman, S., Blair, E., Bradburn, N. and Stocking, C.(1977), Estimates of threatening behavior based on reports of friends, *Public Opinion Quarterly*, Vol. 41, pp. 261-264.
- [5] Warner, S.(1965), Randomized response : A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, Vol. 66, pp. 63-69.

[2003년 1월 접수, 2003년 5월 채택]