

# 통계조사에서의 오차 및 대응방안

최 봉 호<sup>1)</sup>

## 1. 서언

2006년 11월 1일 현재 우리나라에서 공식적으로 작성되고 있는 통계는 모두 704종이나 된다. 작성기관수는 통계청을 포함 모두 154개 기관으로 1개 기관 당 평균 4.6종의 통계를 맡고 있는 셈이다. 704종의 통계를 작성방법별로 살펴보면, 조사통계가 44.3%, 보고통계가 47.4%, 가공통계가 8.2% 수준으로 나타난다. 특히, 통계청의 경우는 조사통계가 43종으로 그 비중이 79.6%나 되고 있다. 그만큼 통계를 생산함에 있어 우리나라에서는 아직도 조사통계가 많은 비중을 차지하고 있음을 알 수 있다.

<표 1> 작성방법별 국가승인 통계 현황

	2004.7.1				2006.11.1			
	계	조사	보고	가공	계	조사	보고	가공
전체	465	237	175	53	704	312	334	58
	100.0	51.0	37.6	11.4	100.0	44.3	47.4	8.3
통계청	52	42	1	9	54	43	1	10
	100.0	80.8	1.9	17.3	100.0	79.6	1.9	18.5

자료: 통계청, 2006년 통계인력 및 예산조사 결과보고서

이와 같이 조사에 의해 만들어지고 있는 통계(추정치)는 항상 오차(표본오차 및 비표본오차)를 수반하게 된다. 즉, 조사에 의한 추정치는 참값(true value)에 100% 일치하지 않을 수 있고, 오차 때문에 참값으로부터 비껴날 수 있는 것이 일반적이라 할 수 있다.

그럼에도 불구하고 조사를 통해 통계(추정치)를 만들어낼 때, 추정치(결과)에만 관심을 두고 그 추정치가 어느 정도의 오차를 포함하고 있는지에 대해서는 큰 관심을

1) 통계청 통계정책국 지역통계과장. 이 글을 작성함에 있어 지역통계과의 직원들로부터 많은 도움이 있었음을 밝힙니다.

기울이지 않고 간과하는 경우가 많다. 특히, 조사의 과정보다 결과만을 중요시하는 풍토 하에서는 오차와 같은 조사의 질에 대한 연구의 중요성이 무시되기 쉽다<sup>2)</sup>.

이런 차원에서 본고에서는 조사통계에서 항상 발생하고 있는 오차(특히, 조사협조 거부로 인한 오차) 현황과 이를 줄일 수 있는 방안에 대해 살펴보고자 한다.

## 2. 통계조사에서의 표본오차

표본오차란 전부를 조사하지 않고 일부만 조사하기 때문에 불가피하게 발생하는 오차로서 표본오차는 표본이 작으면 작을수록 커지게 된다. 이는 표본의 크기가 모집단과 같아지면, 즉, 센서스인 경우 표본오차는 0이 됨을 의미한다. 이를 수식으로 표현하면 다음과 같다.

표본조사를 통한 어떤 추정치에 대한 오차의 한계는  $\pm Z \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{\text{design effect}}$ 로 표현된다. 즉, 어떤 추정치에 대한 표본오차는 보편적으로 세 개의 요소로 구성되어 있다고 할 수 있다.

첫 번째로,  $Z$ 는 신뢰수준 95% 하에서 1.96 이라는 상수 성격의 값을 가진다.

두 번째로,  $s/\sqrt{n}$ 는 표준오차를 의미하는데, 이는 표본의 표준편차와 표본크기와 관련되어 있다. 즉, 표본의 크기가 증가되면 당연히 표준오차가 적어지게 됨을 알 수 있다. 그리고  $s/\sqrt{n}$ 를  $\bar{x}$ 로 나누어준 값은 상대표준오차(relative standard error) 또는 변동계수(coefficient of variation, CV)가 된다. 이 값은 대개 퍼센트로 표시되는데, 0~5%는 매우 우수(excellent), 5~10%는 우수(very good), 10~15%는 좋음(good), 15~25%는 허용가능(acceptable), 25~35%는 유의사항과 함께 사용가능(use with caution), 35% 이상은 공표시 신뢰할 수 없음(too unreliable to publish)과 같은 특성을 보여주고 있다<sup>3)</sup>.

세 번째로, 가구를 대상으로 하는 통계조사의 경우 단순임의 추출방법을 적용하지 않고 집락 추출방법을 사용하게 된다. 이 경우 집락 내에서 가구들 간에 상관관계가 있을 수 있어 단순임의 추출방법보다 표본의 효율성이 적어질 수 있게 된다. 이를 설계효과(design effect, deff)라 하는데, 이는  $\sqrt{\text{design effect}}$  만큼 오차에 영향을 주게 된다. 만약 설계효과가 2라면 약 1.4배만큼, 3이라면 약 1.7배만큼 영향을 미치게 된다고 할 수 있다.

단순임의추출 시 모평균에 대한 신뢰구간 :  $\bar{x} - Z \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + Z \frac{s}{\sqrt{n}}$

2) 통계보고서 발간시 중요변수에 대해 상대표준오차를 제시하지 않은 경우가 상당히 발견되고 있다.

3) Statistics Canada, 2005 Farm Financial Survey, Catalogue no. 21F0008XIB, Jan.2006, p.81.

추정치 및 오차의 한계 :  $\bar{x} \pm Z \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{\text{design effect}}$

신뢰수준 95%하에서의 신뢰계수 :  $Z=1.96$  with 95% confidence level

모평균에 대한 표본 분산 :  $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$

모비율에 대한 표본 분산 :  $(1 - \frac{n}{N})\hat{p}(1-\hat{p})$

표본 크기 :  $n = \text{sample size}$

표준 오차 :  $\frac{S}{\sqrt{n}} = \text{standard error}$

설계 효과 :  $\text{design effect} = 1 + (m-1)\rho$

집락내 상관계수 :  $\rho = 1 - \frac{S_w^2}{S^2}$

상기 식에서 보이듯이 표본조사를 실시하는 경우 표본오차는 항상 발생함을 알 수 있다.

따라서 표본오차를 최소화하기 위해서는 적절한 표본 크기에 대한 연구, 상대표준 오차의 계산, 효율적인 표본추출방법(예: 집락의 크기 등)을 선택하여 설계효과를 적게 하는 것이 최선의 방법임을 알 수 있다.

### 3. 통계조사에서의 비표본오차

#### 가. 조사대상 범위 오차(Coverage error)

조사대상 범위 오차는 모집단에서 개인, 가구, 사업체, 농가 등과 같은 관심대상의 일부가 모집단에서 누락될 때 또는 조사대상 범위 밖의 단위가 모집단에 포함될 때 발생하는 오차를 말한다(inadequacy of sampling frame). 여기서 모집단은 목표모집단(target population)과 조사모집단(frame population)으로 나누어지는데, 조사대상 범위 오차는 목표모집단과 조사모집단간의 차이가 되는 것이다.

예를 들어, 목표모집단이 어떤 지역 내에 거주하는 모든 인구로 정했으나, 조사모집단에서 전화를 가지고 있는 인구로 한정했을 때 발생하는 오차가 바로 조사대상 범위 오차(이 경우 가구에 살고 있으나, 전화가 없는 사람)가 된다. 다른 사례로 우편조사 시 주소지와 실제 거주지가 다른 주소록을 사용할 때 발생하는 오차도 해당된다.

또한 경제활동인구조사에서 15세 이상 인구가 조사대상에 모두 포함되어야 하나 14세 인구가 포함되어 있다든지 또는 15세 이상인구가 빠졌다든지 하는 오차도 조사대상 범위 오차가 되는 것이다(problem of identifying target groups).

상기 사례를 요약하면, 조사대상 범위로차의 원인은 여러 가지가 있을 수 있으나 다음과 같은 원인 때문에 주로 발생되고 있음을 알 수 있다 - i) 불완전한 조사구 요도 및 가구명부, ii) 조사개념의 불명확, iii) 조사원에 의한 고의적 은폐, iv) 조사대상의 한정(기혼부인, 학생 대상 조사 등과 같이), v) 연령의 한정(15세 이상 인구에 대해서만 조사).

## 나. 측정오차(Measurement error)

측정오차는 응답자가 제공한 변수 값과 그 변수의 참값(알려지지 않음) 간에 차이로서 특징될 수 있다. 이와 같은 측정오차는 측정과정이 통제되지 않으면 당연히 발생하게 된다. 왜냐하면 데이터가 수집/측정되는 과정이 질문자와 응답자 간 상호작용에 의해 이루어지기 때문이다. 즉, 사회과학의 통계조사에서는 사람과 사람 간의 상호작용이 측정과정에 개입되어 자연과학의 경우보다 매우 복잡하여 오차 발생 가능성이 매우 높다고 할 수 있다<sup>4)</sup>.

이외 많은 경우 기록된 결과로부터 응답이 도출되기도 하지만, 경우에 따라서는 응답자의 기억에 의존하거나 주관적 판단에 의해 대답되기도 한다. 따라서 측정과정을 자연과학에서와 같이 정형화된 모형으로 만들기는 어렵다<sup>5)</sup>. 사회과학 학자들은 측정오차를 설명하기 위해 타당도와 신뢰도라는 개념을 사용하고 있다. 즉, 통계조사는 타당하며 신뢰도 있게 이루어져야 됨을 강조하고 있다. 여기서 '타당하다'라는 것은 측정결과가 측정 목적에 충실하게 부합되고 있다는 것을 의미하며, '신뢰도가 있다'라는 것은 측정결과가 안정적으로 일관성 있게 나왔다는 것을 의미한다.

일반적으로 통계조사에서의 측정오차는 질문지, 자료수집방법, 면접자, 응답자와 같은 네 가지 때문에 발생하게 된다.

i) 질문지 오차: 질문지는 애매모호하지 않게 응답자와 의사소통이 잘 이루어지도록 설계되어야 한다. 그러나 질문지의 설계(개념의 모호, 문항/조사표의 길이, 문항순서, 개방/폐쇄형 조사표 등), 내용 또는 문구가 조잡하게 되는 경우가 많이 발생된다. 이와 같은 이유로 발생하는 오차를 질문지 오차라 한다.

---

4) 상징적 상호작용론은 '사람들은 일상생활에서 개인적으로 접하게 되는 모든 대상에 의미를 부여하고, 그에 기초해서 생각하고 행동한다' 라고 요약할 수 있다. 또한 상호작용은 상징적이기 때문에 상징물, 언어, 의미 등이 모두 포함되고 있다.

5) 이재창, 통계의 시의성과 정확성, 통계, 제23권 제1호, 대한통계협회, 1997.

ii) 자료수집방법 오차: 자료수집방식(예: 우편, 전화 또는 면접 등)에 따른 오차를 말한다. 응답자들은 면접자 앞에서나 전화 또는 컴퓨터 또는 자기기입 방식에 따라 문항들을 다르게 응답할 수 있다.

iii) 면접자(조사자) 오차: 면접자의 작성에 의존하는 조사에서 면접자는 문항에 대한 응답에 영향을 준다. 이는 면접자가 의도적으로 항목들을 읽지 않거나 응답자를 잘못된 방향으로 이끌 수 있는 다른 정보를 추가함으로써 야기되는 오차를 포함한다. 또한 면접자들은 부적절한 훈련이나 기법들로 인하여 오차를 야기하기도 한다.

이외 면접조사원은 저마다 감수성(sensory acuity)과 판단이 서로 다르며 의견이나 기대 등도 다르다. 따라서 이들은 저마다 응답을 다르게 해석할 수 있다. 그뿐만 아니라 이따금 오기나 다른 란에 기입하는 과오를 범하는 경우도 있으므로 오차가 생길 수 있다. 또한 경우에 따라서는 면접조사원이 고의로 허위기재를 함으로써 오차가 생기게 되는 경우도 있다.

따라서 면접자는 인터뷰하는 동안 중립적인(neutral) 입장을 견지해야 한다.

이 오차와 관련하여, 면접자 오차로 보기는 어려우나 조사를 기획하는 사람은 '특정 질문이 모든 응답자에게 적용될 수 있다'라는 생각과 응답자가 '아주 특정분야에 대한 지식을 가지고 있을 것'이라고 간주하는 경향을 보여준다. 사실, 조사기획자들은, 어떠한 질문은 응답자가 답변을 완성하는데 상당한 작업량이 요구되거나 때로는 영수증이나 가계부를 찾아야 하는데도 불구하고, 곧바로 답변을 할 수 있을 것으로 생각하는 고정관념이 있는 것이다. 이를 고정관념으로 인한 오차라 한다.

iv) 응답자 오차: 응답자들은 자신이 겪은 경험, 지식, 태도 등 때문에 질문지 항목들의 의미를 다르게 해석하여 야기되는 오차를 말한다. 응답자들은 대개 어떤 사항에 대해 과장을 하거나 사회적으로 바람직하다고 여겨지는 답변을 하는 경향을 보이고 있다. 일반적으로 잘 알려진 몇 가지 응답자 오차를 살펴보면 다음과 같은 것들이 있다<sup>6)</sup>.

-망원경 효과(telescoping effect): 이는 어떤 시점에 겪은 사건을 그 사건이 일어난 시점보다 그 이전에 또는 더 최근에 겪었던 것으로 생각하는 경향을 의미한다.

-겸손효과(senor effect): 자기 생각과는 달리 조사자의 기분이나 감정에 맞추는 방향으로 응답하려는 경향을 의미한다. 어떤 경우에는 조사원에게 예의바르고 협조적이며 조사원이 듣기 좋아할 것이라고 생각하는 종류의 응답으로 기우는 경향을 보이고 있다. 즉, 응답자가 조사원의 의도나 입장을 생각해서 거기에 부응하려는 의도로 과잉반응을 보일 수 있는 것이다. 이는 다른 용어로 동정효과(Sympathy effect)라 한다.

-호손효과(hawthorne effect): 표본으로 뽑힌 응답자는 표본으로 뽑혔다는 이유로

---

6) 조사대상이 아닌 부적절한 응답자로부터 정보를 얻기 때문(proxy response) 일어나는 오차도 응답자 오차에 포함된다.

평상시와 다르게 대응하는 경향을 말한다. 예를 들어, 응답할 때 면접자를 가급적 만족시키고자하는 방향으로 응답하는 경향을 들 수 있다.

- 자신향상효과(self-lifting effect): 응답자 자신의 사회적 위치를 현실보다 높게 생각하여 자신의 위치에 해당되는 응답 대신 자기가 이상으로 생각하는 모습에서 응답하게 되는 경향을 말한다. 응답자가 자기자존심에 가장 부합되게 응답하는 경향도 이 범주에 속한다고 할 수 있다.
- 주변 동조효과(conformity effect): 자신의 생각과는 다르게 주위사람들의 영향을 받아 주위 사람을 따라 응답하는 경향을 의미한다.
- 사회적 바람직성(social desirability): 사회적 통념적 가치를 묻는 질문에 통념적 사회규범에 입각하여 응답하는 경향이 있는데 이것을 사회적 바람직성에 의한 응답성향이라고 한다. 사회적으로 문제가 되고 있는 뇌물, 매춘, 청소년흡연 등에 관한 실태조사에는 그러한 응답성향으로 인하여 경험률이 실제보다 과소평가될 가능성이 매우 높게 나타난다. 예를 들어, 미국과 캐나다의 가계조사에서 집계된 음주에 대한 지출이 실제 주류업체에서 판매한 양의 절반정도로 파악되고 있는 것을 들 수 있다. 또한, 사회적 바람직성 때문에 응답자는 흔히 더 많이 소유하고 있다고 과장하거나 실제보다 더 부자라고 주장하거나 자기의 연령을 속이기도 한다.
- 선전효과(bandwagon effect): 선거에 관한 여론조사 등에서 자기 자신의 소신과는 다르게 선두 주자나 동정심이 가는 후보자에게 표를 몰아주어야 한다는 선전에 영향을 받아서 질문에 응답하는 경향을 말한다. 또한 어떤 상품을 사는 사람이 증가할수록, 그 상품에 대한 선호도가 더욱 증대하는 현상도 선전효과에 포함된다.
- 체면치레효과(Ego-threat effect): 유행이나 시대에 뒤떨어질 것을 두려워하여 그릇된 응답을 하는 경향을 의미한다.
- 습관효과(Habit effect): 질문의 내용을 신중하게 생각한 후에 응답하지 않고 단지 습관적으로 예 또는 그렇다는 대답을 되풀이하는 경향을 의미한다.
- 무관심효과(Irrelevance effect): 자신은 조사와 상관이 없다고 생각하여 빨리 끝내기만을 바라고 생각 없이 마음 내키는 대로 아무렇게나 대답하는 경향을 의미한다.
- 절차(프로세스) 상 오차: 조사의 전반적인 절차에 대한 통제가 부족할 때나 조사 대상자가 아닌 사람으로부터 답변을 얻어 낼 때 (proxy response) 오차가 발생할 가능성이 높아지게 된다.

## 다. 무응답(Non-response) 오차

무응답오차는 적정한 조사대상처에서 유용한 정보를 얻지 못하고 실패한 경우 때문 발생하는 오차를 말한다. 무응답오차는 단위무응답과 항목무응답 두 가지 형태로 나누어진다.

단위무응답은 조사대상처인 가구나 사업체에서 정보를 전혀 얻지 못한 경우를 말한다. 자료수집과정에서 최선을 다 한다 해도 어느 정도의 단위무응답은 발생할 것이다. 단위무응답의 이유는 다양하다. 단위무응답은 가구에서 아무도 못 만나거나 (respondent-not-at-home)<sup>7)</sup>, 조사거부(refusal)<sup>8)</sup> 또는 언어장애나 신체장애 등에 의해 발생할 수 있으며, 전화조사에서 조사대상자들이 조사를 직접 거부하거나, 전화 자동응답 장치를 이용하여 조사를 부탁하기도 전에 조사를 거절하기도 한다. 또한 우편조사에서 조사대상자들은 거절의 표시로 조사표를 발송하지 않기도 한다.

항목무응답은 응답자가 조사표 항목을 완벽하게 마무리 짓지 못했거나 얻어진 정보가 유용하지 않을 때 발생한다. 항목무응답도 여러 가지 사유에 의해 발생한다. 항목 누락비율은 특히, 소득과 같이 민감한 사항일 때 가장 높게 나타난다. 때로는 응답자가 질문사항에 대한 정보를 가지고 있지 않을 수도 있으며, 응답자가 다른 사람을 대리할 때 항목무응답이 더 빈번하게 발생한다.

통계청에서 매월 실시하고 있는 조사 중 난이도가 높은 가계조사에서의 가계부 회수율 추이를 <표 2>에서 살펴보면 회수율이 약 80~85% 수준에서 머무르고 있는 것으로 나타나고 있다. 그러나 새로운 표본으로의 개편과 관련 지난 2002년 4/4분기 시험조사 기간 동안 가계부 회수율이 불과 53.7%로 매우 낮게 나타난 적이 있었다. 이는 가계조사의 경우 협조설득에 다소 시간이 걸리고 있음을 뜻한다고 할 수 있겠다. 즉, 조사원과 응답가구 간 유대감(rapport)이 형성되는 최소한의 시간이 필요하다 하겠다.

<표 2> 통계청 가계조사에서의 무응답률 추이

	2002	2002.10~12	2003	2004	2005	2006.1~4
적격가구	5,009	7,613	7,584	7,482	8,692	8,738
무응답률(%)	17.4	46.3	20.6	15.7	16.5	17.0

- 주: 1) 2003년 1월부터 읍/면 지역의 일반가구(234개 조사구)를 포함하여 조사하기 시작하였음. 명칭도 도시가계조사에서 전국가계조사로 변경되었음.  
 2) 2002년 10~12월간에는 2003년 1월부터 시작되는 신 표본지역에 대하여 시험조사를 실시 한 바 있음. 따라서 2002년 10~12월간의 무응답률은 시험조사 결과임.  
 3) 2005년 1월부터 단독(1인)가구를 포함하여 조사하고 있음.

7) 부채자수는 요일, 시간 및 계절 등에 따라 달라지는데, 미국의 경우 대도시에서는 40~50%에 이르고 한다. 이는 재방문(call-backs)이나 이웃집으로의 대체(substituting neighboring dwelling units)에 의해 어느 정도 관리할 수 있다.  
 8) 미국 가계조사에서의 경우 거절율이 25.1%로 나타나고 있다. 이러한 거절은 개성(personality)이나 기분(mood)에 좌우되는 것이므로 우발적이어서 별로 바이어스를 생기게 하지는 않는다고 하나 대체로 고/저 소득층에 많다는 실증자료가 있으므로 오차가 생기지 않는다고 할 수는 없다. 또한 소득에 관한 질문은 거절당하기 쉽다.

위에서 살펴 본 우리나라 가계조사에서의 무응답률 수준(15~20%)은 다른 나라 가계조사의 수준과 비교하여 볼 때 가장 낮은 수준을 나타내 준다(표 3 참조). 폴란드, 헝가리, 영국 및 아일랜드에서는 무응답률이 40%를 상회하는 수준을 보여주고 있다. 미국 및 캐나다의 경우는 각각 25.1% 및 30.8%를 나타내고 있다. 우리나라에서 무응답률이 상대적으로 외국에 비해 적게 나타나는 이유로는 두 가지 때문으로 사료된다. 하나는 외국의 경우 조사원들이 비정규직으로 구성되어 있는 반면에, 우리나라에서는 통계청 조사원들의 대다수가 정규직으로 구성되어 있는 점이 영향을 미쳤다고 할 수 있다. 다른 하나는 외국의 경우 개인주의가 만연되어 있는 사회인데 비해, 우리나라에서는 아직도 통계조사 협조를 '국민의 의무이기 때문'으로 생각하는 경향이 높은 점이 또한 영향을 미쳤다고 할 수 있다.

<표 3> 주요 외국의 가계조사 무응답률(%)

	한국	미국	캐나다	호주	영국	프랑스	아일랜드	폴란드	헝가리
연 도	2005	2001	2004	1998/99	2000/01	1994/95	1999/00	2000	2000
무응답률	16.5	25.1	30.8	22.6	41.2	27.0	45.0	49.2	39.0

자료: 각국 통계청 홈페이지.

다른 한편, 지난 2005년 11월 1일 기준의 인구주택센서스에서 흥미 있는 사례가 <표 4>과 같이 도출되었다. 이 자료는 서울 동대문구청에서 센서스업무에 종사하였던 동사무소 직원 76명으로부터 얻은 설문조사 결과로서 전국을 대표할 수는 없겠으나 나름대로 시사점을 주고 있다.

동 표에서는 조사원들이 가구를 첫 방문했을 때 가구원부재로 만나지 못한 경우는 21.9%나 되었고, 만났으나 응답 거부당한 경우는 7.8%, 노인/유아 등을 만나 조사실패한 경우는 10.2%로 보여주고 있다. 즉, 조사원들이 가구를 첫 방문했을 때 55.1%만이 적절한 응답자를 만나서 조사에 성공했음을 알 수 있다.

<표 4> 2005년 인구주택센서스에서의 가구에 대한 첫 방문 시 조사협조 여부

	만났음	만 난 후 의 결 과			가구원 부재로 조사 실패
		만나서 조사	만났으나 응답 거부로 조사실패	부적합한 가구원을 만나 조사 실패	
100.0	73.1	55.1	7.8	10.2	21.9

자료: 이 자료는 2005년 인구주택센서스 업무에 직접 종사하였던 서울특별시 동대문구 26개 동사무소 주무, 통계담당 및 센서스 총관리자 76명을 대상으로 실시한 설문조사 결과임.



따라서 조사원들은 인구주택센서스 조사기간 동안 조사실패를 조사협조로 전환코자 가구를 여러 번 방문하였었다. 그럼에도 불구하고 전체가구 중 11.1%에 대해서는 가구원 부재 및 면접거부 등으로 직접 조사하지 못한 것으로 분석되고 있다. 이와 같이 조사가 되지 못했던 부분에 대해 조사원들은 결국 통/반장의 협조를 받거나(proxy response), 주민등록을 이용하여 조사표를 작성하였다고 할 수 있다.

이밖에도 통계청에서는 조사원 중 일부(472명)를 대상으로 조사환경에 대한 설문조사를 2006년 5월 기준으로 실시한 적이 있다. 동 설문조사 결과를 <표 5>에 수록하여 놓았다. 동 표에서는 통계청의 조사원들이 경상적인 표본조사를 위해 조사대상처를 방문할 때, 절반 정도는 방문 당일 조사를 완료하고, 나머지는 조사불응 또는 가구원 부재로 인해 조사를 하지 못하고 있음을 보여 주고 있다. 반면, 농어가 경제조사 및 사업체대상 조사의 경우는 면접완료율이 가구대상 조사의 경우보다 다소 높은 68% 수준을 나타내고 있다. 조사 불응이 적게 나타는 이유로는 통계청 조사원들의 적극적인 설득 노력 때문이 아닌가 싶다.

<표 5 > 통계청 경상표본조사에서의 당일 방문 면접완료율

	계	면접 완료	조사 불응	가구원 부재
경제활동 인구조사	100.0	50.6	3.3	46.1
가계조사	100.0	46.9	13.1	40.0
농/어가 경제조사	100.0	68.0	-	32.0
사업체 대상 조사	100.0	68.3	13.0	24.6

상기 사례 이외의 최근 사례로는 서울지방통계청이 2006.9.18~9.29(11일) 간 실시한 마포구지역 사회통계조사를 들 수 있다. 표본규모는 50개 조사구에 분포되어 있는 1,000가구이었으며, 조사원 25명과 지도원 3명이 맡아 실시되었다. 조사원이 당초(initial) 가구를 접촉한 결과를 살펴보면, 응답협조율은 79.0%로 나타났다. 조사구별로는 응답협조가 거의 완벽하게 이루어진 경우는 5개에 불과하였으며, 24개는 70~89%로 나타났다(표 6 참조).

<표 6> 마포구 사회통계조사에서의 응답협조율 조사구 분포

	0~49%	50~69%	70~89%	90~99%	100%
50개 조사구 (평균 79.0%)	1	10	24	10	5

조사대상 가구원은 총 1,944명<sup>9)</sup>이었는데, 이 중 조사에 협조하지 않았거나 조사하기가 불가능하였던 가구원은 6.3%인 122명으로 나타났다. 이 비율은 항상 나올 수 있는 비율로 볼 수 있으며, 조사하기에 쉽지 않은 비율이라 할 수 있다. 따라서 가구를 대상으로 하는 조사에 있어 이 부분의 인구가 누락되지 않도록 각별한 주의가 필요하다 하겠다 (표 7 참조).

<표 7> 마포구 사회통계조사에서의 조사불응/불능 가구원 분포

가구원의 조사 불응/불능 사유	122명
사생활 노출이 싫어서	63명
출장 및 여행 중이어서	27명
장기 부재	16명
장애인	6명
치매 및 노환	6명
병원에 입원 중이기 때문	4명

다른 한편, 사업체를 대상으로 하는 조사에서도 상황은 만만치가 않다. 2005년 기준 도/소매업 및 서비스업 센서스를 앞두고 3개 지역(서울, 부산, 경기 수원) 80개 조사구에 분포된 4,223개 사업체를 대상으로 실시한 시험조사 결과를 살펴보면, 631개 휴/폐업 사업체를 제외하고 순수한 응답 불응률은 3.4%로 나타났다. 지역별로는 서울 지역에서 5.8%로, 부산지역에서 0.8%, 경기도 수원지역에서는 1.1%로 나타났다. 항목 불응률의 경우는 문제가 더욱 심각하여 사업실적을 물어보는 항목에 대해서는 불응률이 39.8%, 사업자등록번호 항목에 대해서는 10.7%가 응답을 하지 않은 것으로 나타났다<sup>10)</sup>.

<표 8> 사업체조사에서의 불응률(2005년 기준)

	불응률 (%)
사업체 기준 - 서울지역	5.8
- 경기도 지역	1.1
항목 기준 - 사업실적 항목	39.8
- 사업자등록번호 항목	10.7

9) 1,944명은 1,000개 조사대상 가구에서 15세미만 가구주 가구(1개)와 15세 미만 연령층에 있는 가구원을 제외한 수치임.

10) 통계청 서비스업통계과, 2005년 기준 도소매업 및 서비스업 총조사 2차 시험조사 결과. 2005.5.

#### 4. 무응답에 대한 통계청의 대응 방법

무응답을 최소화하기 위한 대책은 사전적 (예방)조치와 사후적 조치로 나누어 조치를 취할 수 있다. 사전적 예방조치들 중에서 우편조사와 전화/면접 조사에서 적용될 수 있는 개괄적인 방법들은 조사통계연구회에서 발간한 보고서<sup>11)</sup>에서 찾아볼 수 있다.

##### 가. 사전적 조치

통계청의 경우는 경상적인 여러 가지 표본조사에 대한 협조도를 높이기 위하여 다음과 같은 조치를 취하고 있다.

- 조사담당자 교육
  - 통계교육원을 통한 전문 교육
  - 본청 해당과의 조사별 순회교육 및 소집교육
  - 지방 청/사무소의 자체 조사지침 및 소양 교육
- 응답자와 유대 강화
  - 조사대상처의 경조사 참석
  - 조사대상처 및 조사대상 지역에 일손 돕기
  - 민원업무 대행(등본발급 등)
- 답례품 지급: 가계 5회, 경찰 7회, 농·어가 12회, 사업체 1회, 소비자물가 1회
- 통계홍보
  - 응답자 간담회 개최: 2005년(237회 6,406명), 2006년 6월말 현재(72회 1,808명)
  - 통계청장 명의 인사장 및 감사장 배부
  - 매월 홍보물(생활의 지혜) 배부
- 연동표본제(sample rotation) 도입
  - 경제활동인구조사 및 가계조사의 연동표본제 도입(조사기간이 5년에서 3년으로 단축)

---

11) 조사통계연구회, 무응답 오차, 자유아카데미, 2000, pp.24 ~ 28.

## 나. 사후적 조치

무응답 부분에 대한 사후적 조치를 위하여 통계청에서는 다음과 같은 방법을 적용하고 있다. 농가 및 어가경제조사의 경우는 <표 8>에서 보이듯이 양은 많지 않지만, 조사하기가 가능하지 않거나 불응한 농가 및 어가에 대하여 조사구내에서 특성이 유사한 농가 및 어가로 대체(substitution)를 해 주고 있다.

<표 8> 농가 및 어가경제조사에서의 조사불능 및 조사불응 가구

	2003년		2004년		2005년	
	조사불능	조사불응	조사불능	조사불응	조사불능	조사불응
농가경제조사 (n=3,200)	37	29	4	33	6	29
어가경제조사 (n=1,175)	24	39	18	16	14	11

가계조사에서는 주택이 재개발로 인하여 철거되었거나, 이사를 간 경우 등은 농가 및 어가경제조사와 마찬가지로 특성이 유사한 가구로 대체를 해 주고 있다. 그러나 조사에 불응한 가구(2005년의 경우 16.5%)에 대해서는 대체를 해 주지 않고 무응답 분만큼의 가중치를 만들어 조치하는 방법을 적용하고 있다. 가중치는 가구주의 거주 지역(서울, 광역시, 도), 직업(사무직, 생산직, 무직, 자영업), 및 가구원수 등 세 개의 특성 구조를 분석하여 총 50개 층으로 세분한 후 각각의 층별로 가중치를 만들어 적용함으로써 정확도를 높이고 있다.

활동인구조사에서는 무응답이 많지는 않으나(2005년의 경우 1.7%), 가계조사의 경우와 마찬가지로 가중치를 만들어 적용하고 있다. 가중치는 남녀별 인구 2개, 5세 그룹별 인구 11개, 서울/6개광역시/9개도의 시부/군부등 25개 총 550개 Cell로 추계인구를 만들어 이 추계인구와 조사결과를 대비한 가중치를 만들어 적용하고 있다.

## 5. 장기적 관점에서의 대응 방안

최근 많은 통계작성기관들은 조사통계를 실시함에 있어 가구나 사업체로부터 개인 정보 보호의식 증대, 맞벌이 가구 및 1인 가구 증가, 응답부담 등의 이유로 조사협조를 얻어내는데 많은 어려움을 겪고 있다. 더구나 무응답도 문제지만 부실하게 또는 허위로 조사되는 경우도 상당히 있어 문제의 심각성을 더해주고 있다.

예를 들어 매년 실시되고 있는 xx업 통계조사의 경우 정확히 기재된 조사표가 거의 제출되지 않고 있다. 따라서 사후적으로 대차대조표, 손익계산서, 원가명세서를 다

시 받아, 아니면 전화로 일일이 결산서 내용을 항목별로 다시 물어 기재내용의 오류를 수정하고 있다. 따라서 통계조사 시 협조를 얻어내기 위하여 응답한 사업체나 가구에 대해 세금감면 혜택을 주는 방안과 부실/허위 조사표 제출에 대해서는 벌금부과 등과 같은 강력한 제제가 필요하다는 인식을 가지고 있는 사람들도 있다.

그러나 이와 같이 획기적인 방안도 있겠으나, 일반적인 대응방안으로 다음과 같은 방안이 많이 제시되고 있다 - i) 새로운 통계를 개발하거나 기존의 통계를 개선할 때, 통계조사를 실시하기 보다는 행정(등록, 신고) 자료를 이용하는 방안, ii) 효율적인 조사표 설계, 효율적인 표본 설계로 표본 크기를 줄이는 방안 및 연동표본의 도입, iii) 자료수집에 있어 IT 기술 활용 등 새로운 기법을 도입하는 방안, iv) 응답자와의 커뮤니케이션을 개선하는 방안, v) 통계작성기관 간 유사한 내용의 통계조사 실시 방지 및 기관 간 유사한 내용 자료의 공유 확대 방안 등. 이에 대해 다음에서 보다 자세히 살펴 보고자 한다.

## 가. 행정기록 자료의 활용

캐나다 통계청의 경우 가계조사를 실시할 때 소득세 신고 자료를 활용하고 있으며, 2006년에 실시한 인구주택센서스에서 일부 세무신고 자료를 이용하였다. 연간주기로 실시되고 있는 사업체조사의 경우에는 총 사업수입의 하위 10%를 차지하는 소규모 사업체에 대해서는 세무신고 자료가 이용되고 있으며, 매월 실시되고 있는 사업체조사와 관련하여서도 부가가치세 자료가 이용되고 있다.

덴마크 등 많은 서구 국가의 경우는 인구주택센서스가 거의 100% 주민등록 자료를 기초로 하고 있다. 주민등록부에는 주민의 기본적인 속성뿐만 아니라 인구동태사항 및 이동사항이 계속적으로 기록되기 때문 통계작성의 기본자료 출처로서도 매우 중요한 역할을 할 수 있는 것이다.

물론 주민등록을 기초로 한 인구센서스는 전통적인 인구조사 방식과 비교하여 볼 때 장/단점을 가지고 있다. 장점은 인구센서스를 실시할 때 주민등록부를 활용하게 되면 응답자의 응답 부담을 줄일 수 있을 뿐만 아니라 비용도 상당히 절감할 수 있는 장점이 있다. 반면, 단점은 다음과 같은 것이 있다.

주민등록부에 기록되어 있는 항목이 제한되어 있다는 점이다. 즉, 주민등록부에는 통계적으로 중요한 많은 항목(직업, 산업, 통근/통학 수단, 주택의 특성 등)이 빠져 있어 한계가 있다.

또한 주민등록부는 급속하게 변화하고 있는 경제사회 현상을 즉각적으로 반영하기에 한계가 있다. 예를 들면 직장 이동, 파트타임 취업, 지하 경제 등이 증가하고 있지만 주민등록부로부터 이와 같은 새로운 현상을 파악하기에는 한계가 있는 것이다.

행정자료를 이용한 새로운 통계개발에 관한 최근의 대표적인 국내 사례는 행정자치

부가 토지대장 및 주민등록전산 자료를 이용하여 토지소유에 관한 통계를 들 수 있다. 외국의 사례로는 네델란드 통계청이 각종 행정기록 자료와 통계추정기법을 이용하여 전통적인 인구센서스 대신에 Virtual Population Census를 개발하여 2001년 기준 자료를 2004년 8월에 공표한 사항을 들 수 있다. 즉, 네델란드 통계청에서는 1981년 기준 인구센서스에서의 무응답률이 약 25%이었던 점을 감안하여 1991년 기준의 인구센서스는 실시하지 않았고, 2001년 기준으로 '가상적인 인구센서스' 결과를 만들어 낸 것이다<sup>12)</sup>.

## 나. 조사원과 응답자간 상호작용 연구 강화 필요

미국 미시간 대학교 조사연구센터의 Gloves와 Couper<sup>13)</sup>는 통계조사 시 가구와의 관계에 있어 조사원이 '가구를 처음으로 접촉할 때(the likelihood of contacting)'와 접촉한 후에 있어서 협조를 얻어내는 측면(the likelihood of cooperation)' 두 가지로 구분하여 특성을 살펴보아야 한다고 주장하고 있다.

이 중 조사원이 가구를 처음으로 접촉할 때 접촉이 제대로 잘 이루어졌는지에 영향을 미치는 요인으로 다음과 같은 것들을 제시하고 있다. 이와 같은 네 가지 특성은 조사원의 자질과 조사원이 가구를 방문하는 횟수 및 방문시간대와 상호 맞물려 가구와의 접촉이 제대로 잘 이루어졌는지가 결정된다고 할 수 있다.

- 사회 환경적 특성
- 인구 노령화, 응답자의 건강 등 인구학적 특성
- 가구에의 물리적 접근가능성
- 응답자가 집에 머무는 패턴

다른 한편, Groves와 Couper는 조사원이 가구와 접촉을 한 후 응답자로부터 협조를 얻어냄에 있어 미치는 요인으로는 네 가지 요인(사회/환경적 배경, 조사원의 특성, 조사 설계 측면, 응답자 특성)을 제시하고 있다. 이 네 가지 요인이 서로 작용하여 응답자가 조사에 협조할지 아니면 불응할 지 결정을 내리는데 영향을 미치게 되는 것이라 할 수 있다.

즉, 조사원이 통계조사를 위해 가구를 방문할 때 응답자의 반응은 조사원과 짧은 시간(1~5분)동안 부딪히면서 조사에 협조할지 아니면 불응할 지를 결정을 하게 되는 것이다. 이 과정에서 응답자는 자기의 기억, 과거의 경험, 지식, 믿음, 태도 등과

12) The Dutch Virtual Census of 2001: Interview with Eric Schulte Nordholt, ISI Newsletter, Vol.28, No.3, 2004.

13) Groves, R. M. and Couper, M. P., Nonresponse in Household Interview Surveys, 1998, Wiley, New York.

주변상황을 모두 고려하여 조사협조요청에 대응을 하는 것이다. 반면, 조사원들은 여러 가지의 설득 전략을 가지고, 응답자의 말이나 행동이 무엇을 의미하는지 파악코자 노력하면서 응답자의 협조를 얻어내고자 노력하고 있다고 할 수 있다<sup>14)</sup>.

이와 같은 조사 참여 이론체계에서는 조사원의 행동은 표준화되어야 한다는 전통적인 생각이 부정되는 것이다. 실제적으로 이 이론의 핵심은 조사원들은 면접조사 시 응답자의 특성(수준)에 적절하게 맞추어 접근해야 한다는 것이라 할 수 있다. 이와 같이 이제까지 현장에서의 조사과정에 대한 이론적 연구가 소홀히 취급되었던 같다. 이런 차원에서 현장에서의 조사환경, 조사원과 응답자 간의 상호작용관계 등에 대한 연구가 강화되어야 하겠다.

통계청이 실시하고 있는 통계조사의 경우 아직까지 위에서 살펴 본 여러 가지 특성별로 응답자의 협조도 분석을 구체적으로 시도하지 못하였다. 대신 경제활동인구조사, 가계조사, 농가경제조사 및 어가경제조사별로 협조를 해 주고 있는 사유 다섯 가지를 2006년 2월 시준으로 설문조사한 바 있는데 그 결과가 <표 9>에 제시되었다.

동 표에서 경제활동인구조사에 대한 응답가구는 ‘국민의 의무이기 때문에’ 응답한다는 비율이 67.2%로 가장 많이 나왔다. 경제활동인구조사와 가계조사를 같이 응답하고 있는 가구의 경우는 앞에 조사와 마찬가지로 ‘국민의 의무이기 때문에’가 50.9%로 가장 높게 나타났다. 농가경제조사의 경우와 어가경제조사의 경우도 ‘국민의 의무이기 때문에’가 가장 높게 나타났으나, ‘조사원과의 유대감’ 때문에 답한다는 비율도 각 23.9%, 28.2%나 되고 있다.

<표 9> 통계청 실시 통계조사에 대한 응답사유(2006년 2월 기준)

(단위: %)

구 분	국민의 의무이기 때문	가계수지 파악을 위하여	답례품 때문	조사원 유대감	기타
전체	60.4	6.1	15.0	16.2	2.3
경찰	67.2	1.2	15.6	13.4	2.6
가계	50.9	14.3	15.4	16.7	2.7
농가	51.5	14.4	9.5	23.8	0.8
어가	43.0	12.5	15.4	28.2	0.9

자료: 2006년 3월 기준 통계청의 경제활동인구조사 대상가구 중 6,516 가구, 농가경제조사 대상가구 중 640 농가, 어가경제조사 대상가구 중 552 어가로부터의 설문조사 결과임.

14) Groves, Robert and Mick Couper, Contact-Level Influences on Cooperation in Face-to-Face Surveys, Journal of Official Statistics, Vol.12, No.1, 1996.

## 다. 통계기술적인 측면

자료수집방법도 측정오차 및 무응답문제와 밀접히 관련되어 있는 것으로 알려져 있다. 따라서 전통적인 조사원에 의한 면접조사방법 이외 다양한 자료수집방법에 대한 검토가 필요하다. 우편조사, 전화조사, Fax에 의한 조사, 인터넷 조사 등과 같은 방법에 대해서 적용가능성을 검토할 시점이라 하겠다.

일본 수리통계연구소 주관으로 1953년부터 매5년 주기로 실시하고 있는 “국민성 조사”의 경우 이제까지 매 조사마다 조사내용을 종전과 같게, 자료수집방법을 조사원에 의한 면접 방법을 계속 고수해왔다. 그러나 무응답률이 2003년 기준 조사 시에 56.0%까지 저하된 것과 관련하여 전통적인 조사원에 의한 면접방법을 응답자가 회망하는 우편조사, 전화조사 및 인터넷 조사방법으로 전환코자 많은 연구를 하고 있다.

또한 무응답에 대한 연구(발생원인, 처리방법 등)도 이제까지 소홀히 되었다고 할 수 있다. 이유로는 조사의 과정보다 결과만을 중요시하는 풍토 하에서 오차와 같은 조사의 질에 대한 연구의 중요성이 무시되었다고 할 수 있다. 특히, 매월 또는 분기별 표본조사 결과로부터 계산되는 증감률 자료에는 무응답이 미치는 영향이 1회성 조사 때보다도 크다고 할 수 있으나, 이에 대한 연구가 거의 이루어지지 않고 있다.

따라서 조사에 협조를 하지 않은 사람들로 부터 자료를 얻어내기 쉽지는 않겠으나, 무응답자들에 대한 기본적인 인구특성 및 조사불응 사유를 파악코자 하는 노력이 필요하다 하겠다. 아울러 무응답자를 인구센서스에서 조사된 자료와 개별적으로 연결하여 분석하는 노력도 시도할 필요가 있다 (Census Linked Study). 이외 무응답에 대한 가중치를 부여할 때 무응답이 발생한 지역을 별도의 여러 가지 소지역자료를 비교/분석하여 그 지역특성에 맞게 가중치를 만들어 주는 방법에 대한 연구도 필요하다 하겠다.



## <참고 문헌>

- 김달호, 김남희 (1998), 신뢰성있는 지역통계 추정을 위한 제언, *Journal of the Korean Data & Information Science Society*, Vol. 9, No. 2.
- 김병진, 소재진 (2006), *사회복지 조사방법론*, 법문사.
- 박홍래 (2005), *통계조사론*, 개정판, 영지문화사.
- 샘플링 아카데미 (2005), *표본조사 입문*, 개정판, 자유아카데미.
- 조병희, 설문지 작성기획, 2006년도 한국조사연구학회 워크숍 발표 파워포인트 자료.
- 조사통계연구회 (2000), *무응답 오차*, 자유아카데미.
- 조사통계연구회 (2005), *조사방법의 이해*, 개정판, 교우사.
- 정구현 (2006), *가계조사 추정방법 변경안: 무응답 처리방법 및 분산추정식*, 통계청 내부자료.
- 최봉호 (2006), *통계조사시 조사원과 응답자 간 상호작용에 관한 연구*, 한국자료분석학회 2006년도 추계 학술논문 발표대회 발표자료, 2006.10.21.
- Federal Committee on Statistical Policy(2001), *Measuring and Reporting Sources of Error in Surveys*, Statistical Policy Working Paper 31, Office of Management and Budget.
- Gloves, Robert (1989), *Survey Errors and Survey Costs*, Wiley.
- Gloves, Robert and Mick Couper (1998), *Non-response in Household Interview Surveys*, Wiley.
- Groves, Robert and Mick Couper (1996), *Contact-Level Influences on Cooperation in Face-to-Face Surveys*, *Journal of Official Statistics*, Vol.12, No.1.
- Haraldsen, Gustav (2004), *Identifying and Reducing Response Burdens in Internet Business Surveys*, *Journal of Official Statistics*, Vol.20, NO.2, pp. 393-410.
- International Statistical Insitute (2004), *The Dutch Virtual Census of 2001: Interview with Eric Schulte Nordholt*, *Newsletter*, Vol.28, No.3.
- Kish, Leslie (1965), *Survey Sampling*, Wiley.
- Lehtonen, Risto and Erkki Pahkinen (2004), *Practical Methods for Design and Analysis of Complex Surveys*, 2nd Edition, Wiley, 2004.
- McDonnell, Paul (2005), *Weighting the National Travel Survey*, Paper presented at 9th Meeting of the National Statistics Methodology Advisory Committee, UK Office for National Statistics, London.
- Statistics Canada (2006), *2005 Farm Financial Survey*, Catalogue no. 21F008XIB.
- Ward, Denis (2003), *OECD Summary of National Practices to Reduce Cost and Burden to Respondents and NSOs*, Paper presented at OECD Short-term Economic Statistics Expert Group meeting on 26~27 June 2003.
- Wolter, Kirk (1985), *Introduction to Variance Estimation*, Springer.