

## 조사통계의 과거, 현재, 미래

신민용<sup>1)</sup>, 이흥철<sup>2)</sup>

### 1. 서론

조사통계의 역할은 통계 이용자들이 요구하는 통계자료들을 생산, 제공하는 것이라고 할 수 있다. 21세기에 들어 지식, 정보화 사회로의 사회변화가 급진전되면서 통계기능도 크게 달라져 오늘날의 통계는 정부, 기업, 개인의 자유로운 판단과 의사결정을 뒷받침하는 사회의 정보인프라의 핵심적인 구성요소의 하나로 인식되고 있다. 이와같이 통계조사는 그 중요성이 높아가고, 규모도 급속히 확대되어 가고 있다. 국가기관에서 실시하고 있는 각종 통계 조사, 조사업체에서 마케팅 조사, 사회단체에서 시행하는 사회조사, 공업분야에서 실시하는 품질조사가 있는데, 그 외에도 셀 수 없이 많이 표본조사(sample survey)가 이루어 지고 있다. 조사통계 업무의 발전 방안으로 과거와 현재를 돌아 보면서 미래에 유의하여야 할 문제점을 지적하고자 한다. 우리가 논의하고자 하는 주제는 크게 나누어, 자료의 수집방법, 표본조사 실시방법, 자료의 분석 방법이다. 간략히 말하여 자료의 수집방법은,

- (1) 표본조사                      (2) 직접측정(direct measurement)
- (3) 관찰(observation)        (4) 이차적 정보(secondary information)

등이 있다.

표본조사 실시 방법으로는,

- (1) 우편 조사      (2) 면접 조사      (3) 전화 조사      (4) 인터넷 조사

등이 있다.

앞으로 많이 활용되어야 할 자료의 분석 방법으로는,

- (1) 무응답에 대한 분석
- (2) 콤플렉스 서베이(complex survey)에서 분산 추정 (variance estimation)
- (3) 콤플렉스 서베이(complex survey)에서 범주형 자료분석
- (4) 콤플렉스 서베이(complex survey)에서 회귀 분석
- (5) 소지역 추정( small-area estimation)

등으로 그 외에도 많은 분석 방법이 있다.

### 2. 조사통계의 현황과 전망

#### 2.1. 조사통계의 어제

조사업계를 하나의 산업이라는 측면에서 볼 때, 한국 조사산업은 크게 네 시기로 나누어 볼 수 있다. 우선은 1960년대 말부터 1970년대까지 조사업계의 태동기라고 볼 수 있는 시기, 1980년대의 정착기, 1990년대 확장기, 그리고 1998년 이후 재구조화기로 볼 수 있다.

한국갤럽이 국내에서 공식적인 여론조사를 시작한 것은 1980년 한국일보의 의뢰로 조사한 전

1) 한국외국어 대학교, 교수

2) 이흥철 (NI Korea 대표)

국민 여론조사가 최초라고 볼 수 있다. 물론 1979년에 일본 정부의 의뢰로 5개국이 공동으로 조사한 아동과 어머니를 대상으로 한 국제비교조사가 있었지만, 국내 발표를 목적으로 한 여론 조사는 1980년도 한국일보와 함께 한 여론조사가 최초였다. 그 후 한국갤럽은 국제기구와 함께 200여 회의 국제 여론조사를 실시했으며, 이 분야에서는 국내에서 한국갤럽만의 독보적인 기록이라고 볼 수 있다.

한국리서치는 1990년대 말까지는 여론조사는 수행하지 않고, 순수한 마케팅 전문기관으로 성장하였다. 그래서 1980년대 초에 마케팅 조사 분야에 다양한 조사기법과 분석기법 등을 정착시키며, 현재까지 기업의 의뢰를 받고 사안마다 조사를 진행하는 AdHoc 마케팅 조사분야에서 국내 선도적인 회사로 성장하였다.

1980년 세계 최대의 조사기관 그룹인 A.C.Nielsen이 한국지사를 설립하여 조사를 위한 준비 작업을 하다가 1982년부터 정기적인 소매점 패널조사를 통한 소매점지표 서비스를 개시하였다. 1980년대 후반부터는 A.C.Nielsen은 소매점 패널을 통한 유통조사분야에서는 거의 국내 시장을 독점하며, 국내에서 가장 규모가 큰 조사기관으로 성장해 갔다.

1968년부터 A.C.Nielsen 한국지사가 설립하게 된 1982년까지 국내 조사시장은 거의 불모지나 다름이 없었다. 한국 갤럽, 한국리서치, 그리고 A.C.Nielsen 세 회사 모두 1980년대 중반까지는 재정적으로 많은 어려움을 겪었다. 그 이유로 우선은 소비자를 대상으로 하는 조사 자체가 국내 기업에는 생소했을 뿐만 아니라, 당시의 국내 기업의 마케팅은 유통망 중심의 판촉이 주류를 이루었기 때문에 소비자 대상의 시장조사에 대한 필요성을 거의 느끼지 못했다.

## 2.2 조사통계의 오늘

국내 조사산업은 1968년 처음으로 조사기관이 시작된 이래 30여 년 동안 꾸준히 성장해 왔다. 상업적인 조사기관의 수는 50여개 업체(온라인 조사기관 제외)를 상회하는 것으로 추산되며, 이들 조사기관의 총 매출규모는 2000년 기준으로 1000억원을 상회하는 것으로 추산된다. 현재 조사업계 종사하는 인원은 대략 1,000명정도이고, 그 중 연구원이 약 400명, 실사지도원이 약 300명, 사무보조원 와 전산요원이 약 300명 정도인 것으로 추산된다. 전체 매출 중 한국마케팅 여론조사협회에 가입한 회원사들의 매출이 전체의 약 80%를 차지하며, 한국마케팅여론조사협회에 가입하지 않은 비회원사의 매출이 전체의 약 15%, 그리고 광고 대행사에서 조사업계에 외주를 주지 않고 자체적으로 수행하는 조사매출이 약 5%를 차지하는 것으로 추산된다. 1980년에 추산된 조사업계의 매출이 약 5억원인 점을 감안하면 20년 동안 약 200배의 성장을 해온 산업이며, 최근 몇 년 사이에도 IMF를 이듬 해인 1998년을 제외하면 매년 20-30%의 성장을 해왔다. 1999년도를 기준으로 볼 때, 국내 조사업계의 매출은 아태지역 조사시장의 약 4%를 차지해, 일본 중국 다음으로 외형이 큰 시장으로 자리잡았다.

국내에서 대표적인 자료분석서비스 분야로 자리잡은 조사업계는 시장조사나 여론조사에 대한 필요성이 거의 없었던 척박한 환경에서 지난 30여년동안 많은 어려움을 겪으며 성장해 왔으며, 앞으로도 유망한 직종중의 하나로 지적되고 있다.

## 2.3 조사통계의 내일

2000년대에 접어들면서 여러가지 다양한 조사기법과 상품들이 만들어 지고 있다. 특히 온라인 등과 같은 새로운 채널을 이용한 조사기법의 개발들이 크게 요구되는 시점에 있다. 이런 제반 환경을 토대로, 향후 전망 및 과제를 언급해 보고자 한다.

### 2.3.1 Access 패널의 활성화

앞서 언급했듯이, 국내에서 가구방문에 의한 일대일 개별면접은 구체적인 샘플링프레임도 없을 뿐만 아니라, 면접원이 만들어내는 오차가 심각한 수준이다. 하지만 개별면접을 통해 얻어낼 수

있는 정보는 다른 접근방법에 비해 많은 장점을 갖고 있다. 그런 점에서 새로운 접근 방법에 대한 모색을 조사회사마다 해오고 있으며, 그 중 일반적인 고려가 대규모 패널(Access Panel)에 의한 조사이다. 그 동안 개별 조사회사들이 축적된 조사자료를 토대로 대규모의 패널을 구축하고 이들을 대상으로 자사를 하는 방법을 말한다. 이런 방식의 조사는 1990대 이후 구미에서 광범위하게 활용되고 있는 접근 방법으로, 그 동안 조사관련 세미나에서 그에 대한 많은 이론적인 검토가 있어 왔다. 국내에서도 이제 이런 패널조사에 대한 필요성과 불가피성을 조사 의뢰인들도 공유를 하여, 보다 정교한 조사가 이루어 질 수 있도록 공동의 노력이 필요하리라 사료된다.

### 2.3.2 인터넷 조사 방법의 활용

국내 인터넷 사용인구는 13세 이상에서는 모집단의 3/4에 육박하고 있다. 이런 추세를 감안하면, 현재 인터넷을 활용한 조사에 대해 보다 진지한 방법론적 검토가 있어야 할 것이다. 그런 측면에서 2000년 미국대선에서 Harris interactive는 온라인 조사를 통해 가장 정확한 결과 예측을 하였다는 점, 그러기 위해 Harris interactive는 나름대로 모집단의 특성 추정을 위한 많은 노력들을 해 왔다는 점이 시사하는 바가 크다. 이를 위해서 조사업계와 학계간에 모수치 추정을 위한 다양한 노력들이 필요하리라 보여진다.

또한 인터넷을 활용한 조사의 경우, 단순한 설문조사의 부가적인 채널 이상의 의미를 갖고 있다. 예를 들어, 인터넷이라는 채널이 갖고 있는 상호작용성과 하이퍼링크 기능, 그리고 멀티미디어 제공환경과 사용자에 대한 개인화 가능성을 고려하면, 인터넷이야말로 보다 정교한 실험설계를 통한 측정의 새로운 접근방법으로서의 장으로 제공해 준다. 더욱이 인터넷이 갖고 있는 채널의 특성상 측정과 실제 행동을 연계시켜서 분석할 수 있는 가장 훌륭한 자료원으로서의 잠재적인 가치도 매우 크다.

이외에도 앞서 언급한 Web TV시대를 겨냥한 새로운 시청률 측정시스템의 개발 등 정보기술의 개발과 더불어 조사업계가 대응하기 위한 많은 과제들이 속속 생겨나고 있다. 그런 점에서 이에 대한 조사업계와 학계의 노력, 그리고 정부 및 관련기관의 광범위한 지원이 필요하리라 보여진다.

### 2.3.3 정부 공공통계의 활용

조사를 위한 이차 자료분석을 위해 정부 공공통계를 광범위하게 활용하고 있다. 국내에서 인구센서스 자료를 가장 빈번하게 그리고 광범위하게 사용하고 있는 분야는 아마도 조사업계일 것이다. 비단 표본 설계뿐 아니라, 다양한 업무 과정에서 인구센서스 자료는 조사회사의 필수적인 자료층의 하나로 활용되고 있다. 그럼에도 불구하고 조사업계에서 활용할 수 있는 형태의 자료가 아닌 것이 많아서 많은 아쉬움이 있다. 샘플링프레임으로 활용하기에는 인구센서스 자료가 제한적인 점이 특히 그렇다. 물론 이에 대해서는 조사업계가 보다 적극적으로 이를 활용할 수 있는 방안들을 모색했어야 하겠지만, 통계청에서도 개인의 정보를 보호할 수 있는 전체 위에서, 보다 정교한 조사를 위해 공인된 조사기관(예를 들어 한국마케팅여론조사협회의 가입사)에 구체적인 샘플링 프레임을 제공해 주는 방안이 검토되어야 할 것이다.

또한 국내 공공기관의 통계들에 대해서도 국내 기업들이 보다 정교한 마케팅 전략 수립을 위한 정보분석의 차원에서 세분화시켜 부가가치를 높일 수 있는 방안에 대한 고려도 필요하리라 보여진다. 이를 위해서는 조사기관과 통계청간의 보다 빈번하고 심도깊은 논의의 장이 마련되어야 할 것이다.

### 2.3.4 조사자료의 공개와 활용

언론에 공표되는 원자료에 대한 공개는 조사의 신뢰도 향상을 위해서는 필수적이다. 원자료의 공개를 통해 조사결과에 대해 재검증을 할 수 있을 뿐만 아니라, 만약 특정한 조사가 문제가 된다면 어디가 문제인지를 설문지와 원자료, 그리고 자료처리과정에 대한 상세한 분석이 가능하기 때문이다. 보다 긍정적인 측면에서는 학계나 조사에 관심이 있는 일반인, 혹은 다른 조사 연구자들이 자료에 대한 재분석을 통해 보다 풍부한 정보를 얻어낼 수도 있고, 유사한 여러 조사를 비

교하고 분석할 수 있는 기회의 제공을 통해서 조사 분석의 질적 수준을 향상시킬 수도 있을 것이다. 그러기 위해서는 그런 필요성을 충분히 만들어 낼 수 있는 학계의 노력도 필요할 뿐 아니라, 조사업계와 언론/방송사와 같은 조사 의뢰인의 노력과 참여가 필요할 것이다. 학계에서는 국내의 현실적인 조사여건에 대한 이해를 바탕으로 보다 조사의 품질을 높일 수 있는 현실적인 분석노력과 이에 대한 논의의 장을 조사업계와 공동으로 만들어가야 할 것이다.

### 3. 조사통계의 발전을 위한 통계기법

#### 3.1 자료의 수집 및 표본조사방법

표본조사에서 응답자가 응답 자체를 안하거나 어떤 질문 항목에 대하여 응답을 하지 않는 경우에 무응답 문제는 표본설계나 분석에 큰 영향을 준다. 무응답을 미리막고, 무응답을 측정하고 그 영향을 줄이는 것은 매우 중요한 일이다.

무응답을 다루는 가장 좋은 방법은 무응답이 일어나지 않게 미리 예방 수단을 쓰는 것이다. 무응답자는 응답자들과 본질적으로 다를 수 있다. 만일 무응답을 무시할 수 없다면 응답자들만을 근거로 하는 추론은 심각한 결점이 될 수 있다.

##### 3.1.1 비표본오차를 줄이기 위한 조사 설계

잘못된 표본설계의 공통 특징은 표본설계를 하는 시간과 무응답을 추적 조사하는데 드는 시간이 부족하다는 것이다. 조사를 처음하는 많은 사람들은 단순히 자료 수집과정에서 야기될 수 있는 잠재적인 문제들을 고려하지 않고 바로 자료수집부터 시작한다. 보통 우편 설문지를 발송하고 회수된 자료를 분석한다. 이러한 조사는 낮은 응답률을 보인다.

#### 3.2 자료의 분석 방법

##### 3.2.1 무응답에 대한 자료 분석

###### (1) 가중-크래스의 형성

가중 크래스(weighting-class) 방법은 비표본 오차를 보정하려는 방법이다. 표본추출된 모든 원소에 대하여 알려진 변수들은 가중-크래스들을 형성하는 데 사용되고, 같은 가중-크래스 내에서는 무응답자와 응답자가 유사할 것이라고 기대한다.

가중-조정 크래스들은 층으로 간주하여 형성되어야 한다. 크래스들은 각 크래스 내에 있는 원소들이 관심있는 주요 변수들에 관해서 가능한 유사하고, 각 크래스간의 응답율이 크래스마다 다르게 나타나도록 크래스가 형성되어야 한다.

###### (2) 사후층화 (poststratification)

사후층화는 가중-크래스와 유사하지만  $h$  사후층의 크기가  $N_h$  일 때  $N_h$ 가 가중치를 조정한다.

###### (3) 대체법 (imputation)

조사에서 결측 항목들은 여러 가지 이유로 나타난다. 대체법은 결측 항목에 대체값을 할당한다. 대체법에는 다음과 같은 것이 있다.

###### [a] 연역적 대체 (deductive imputation)

반복적인 조사에서 사용될 수 있다.

###### [b] 칸 평균 대체 (cell mean imputation)

어떤 칸에 있는 응답 개체들에 대한 평균값으로 대치한다.

###### [c] 핫덱 대체법

같은 클래스(class)에 속하는 개체 값이 대치된다.

**[d] 회귀 대체법**

회귀 분석을 이용하여 결측값을 예측하는 방법이다.

**[e] Cold-deck 대체법**

대치되는 값들이 과거의 조사나 다른 정보로부터 얻어진다.

**[f] 교체 (substitution)**

조사자들은 필드(field)에 있는 동안 대체를 선택할 수 있다.

**[g] 다중 대체법 (multiple imputation)**

각 결측값이 2번 이상 대치된다.

**3.2.2 분산 추정 (variance estimation)**

모평균과 모총계는 가중(weight)을 써서 쉽게 추정된다. 분산을 추정하는 것은 더 복잡하다. 더 나아가 콤플렉스 서베이에서 분산을 구하는 공식이 없는 경우도 있다. 콤플렉스 서베이에서 분산을 구하는 방법은 다음과 같다.

① 선형화( Linearization) 방법

비선형함수를 테일러급수(Taylor series)로 전개하여 선형으로 근사시킨 후 분산추정을 한다.

② 랜덤 그룹(group) 과 재표본추출(resampling) 방법

SRS로 크기 n인 표본이 추출되었다면 크기가 n/R인 R개의 그룹으로 나눈다.

③ BRR(balanced repeated replication)

BRR은 반복절반표본들(replicated half-samples)을 사용하여 분산을 추정한다. 여기서 반복절반표본들은 균형(balanced)되어 있다.

④ 잭나이프(jackknife)

BRR처럼 랜덤 그룹 방법을 확장한 것으로 반복 그룹의 중복을 허용한다. 분산추정과 신뢰구간을 구하는 데 이용한다.

⑤ 붓스트랩(Bootstrap)

표본을 모집단으로 간주하여 재표본추출(resampling)하는 방법이다. 이 방법은 신뢰구간을 직접 구하는 데 적합하다.

⑥ GVF(generalized variance function)

GVF는 많은 조사에서 표본오차를 계산하기 위하여 제공된다. GVF는 연간 발행물의 생산시에 시간을 줄이고 생산의 속도를 빠르게 한다.

**3.2.3 범주형 자료 분석**

표본이 자체-가중(self-weighting)이 아닌 한 칸 빈도수(cell-frequencies)는 범주들의 상대 뜻을 의미하지 않는다. sampling weight가 칸 비율을 추정하는 데 사용된다. 만일 표본이 자체-가중이라면 관찰치만으로 칸 비율을 추정한다.

**3.2.4 회귀분석**

콤플렉스 조사에서 회귀분석을 시행하는 데는 SAS나 SPSS 통계 프로그램을 사용할수 있다. 서로 다른 추출확률을 가질 수 있다. 추출확률이 반응변수와 관계가 있다면, 이러한 서로 다른 확률을 고려하지 않으면 회귀계수를 추정하는 데 바이어스(bias)가 생길수 있다. 집락추출에서도 SAS나 SPSS로 구한 표준오차가 틀릴 수 있다.

**3.2.5 소지역 추정**

소지역 추정방법은 첫째, 해당지역에서 관찰된 표본자료를 이용하여 추정값을 구하는 직접 추정법(DE)과 둘째, 해당지역에서 추출된 표본의 관찰 자료와 이와 유사한 다른 소지역들의 관찰 자

료를 이용하여 추정값을 구하는 간접추정법(IE)이 있다.

#### 4. 결론

조사통계의 발전 방안으로는 자료의 수집을 위한 표본설계에서 분석에 이르기까지 각 단계의 자료처리를 통계적이론에 맞게 하여야 한다. 최근에 이르러 통계적인 기법이 컴퓨터의 발전과 더불어 수준이 높아져 이를 잘 적용하여야 질 높은 통계분석을 할수 있다. 조사통계의 발전을 위하여 몇가지 제안을 내고자 한다.

- (1) 표본설계, 자료수집, 자료분석에 이르기 까지 어느 한 단계라도 완벽하지 않으면 잘못된 결과가 나온다.
- (2) 여러기관에서 조사하다가 보면 중복조사가 될 수가 있는데 이를 통합할 장치가 필요하다.
- (3) 사이버 공간에서의 조사 기법에 대한 연구의 필요성이 높아가고 있다.
- (4) 무응답에 대한 자료분석의 필요성이 높아가고 있다.
- (5) 자료수집후에 내용심사(editing)와 통계적 분석기법의 중요성이 높아가고 있다.
- (6) 통계수요를 조사통계만으로 충족시키는 것은 어렵다 행정자료를 이용하여 통계를 작성하는 시스템을 구축해야 한다.
- (7) 조사자료 및 조사방법의 공개는 조사의 신뢰도 향상을 위해서는 필수적이다.

#### 참고문헌

1. 이흥철 (1996), 여론조사 신뢰도 제고방안. 한국마케팅여론조사협회.
2. 김용한 (1999). 한국리서치산업발달사. (사)한국광고인협회 편, 한국광고산업발달사
3. 표본설계(2001), 신민용, 이상은, 교우사
4. 조사연구(2000), 한국조사 연구학회
5. 조사연구의 방법론적 쟁점(2000). 한국조사연구학회
6. Rohr.S.L.(1999). Sampling: Design and Analysis. Duxbury Press.
7. Kalton.G.(2000). Development in survey research in the past 25 years. Survey Methodology. Vol.26. No.1
6. ESOMAR (2000). Esomar annual study on the market research industry 1999.