

다중 추출틀 조사기법을 적용한 인터넷 조사와 전화조사의 사례연구

김 영 권*, 이 계 오**, 김 주 성***, 박무익****

<요 약>

현재 국내의 여론조사에서는 전화조사가 주로 이용되고 있으며, 최근 들어 인터넷조사를 활용하는 방법들이 연구되고 있다. 그러나 인터넷조사의 경우 모집단 구조가 편향된 특성을 가지고 있으므로 이에 대한 보완 수정없이 결과의 직접적인 이용은 많은 비표본오차를 발생시킬 수 있다. 그러므로 본 연구에서는 사회여론조사를 수행함에 있어서 편향모집단에서의 인터넷조사로 부터의 추정값과 전화조사를 통한 추정값의 적절한 결합추정방법을 제안하고, 지난 4.13 총선의 조사자료를 이용하여 이들 결합추정의 효율성을 검토하였다.

I. 서 론

여론조사는 많은 나라에서 특정정당의 지지도를 알아보기 위해서 또는 사회·정치·경제적인 문제들에 대한 대중적인 여론을 측정하기 위하여 여론조사기관이나 정부기관 등에 의해서 정기적으로 또는 특정한 시기에 행해져 왔다. 지난 4·13 총선에서도 유권자의 후보자에 대한 지지도 및 각 정당별 지지도를 측정하기 위하여 많은 여론조사 기관을 통하여 각

* 충북대학교대학원 전자계산학과 박사과정, (361-763) 청주시 흥덕구 개신동 산48
(Tel) 043-261-2255, (Fax) 043-273-5928, purunsol@chollian.net

** 공군사관학교 전산통계학과, 충북 청원군 남일면 쌍수리 (Tel) 043-290-6441
kayolee@orgio.net

***충북대학교 통계학과, (361-763) 청주시 흥덕구 개신동 산48
(Tel) 043-261-2256, (Fax) 043-273-5928, kimjs@cbucc.chungbuk.ac.kr

****한국갤럽조사연구소 소장, (110-054) 서울특별시 종로구 사직동 208번지
(Tel) 02-3702-2100, (Fax) 02-3702-2121, mipark@gallup.co.kr

정당이나 언론매체에서 선거여론조사를 실시하였고 언론매체를 통하여 발표되었다. 그러나 일부에서는 이들 조사결과를 신뢰할 수 없다는 의견을 제기하면서 다음과 같은 사실들을 근거로 내세우고 있다. 첫째 비슷한 시기에 조사된 여론조사의 결과가 발표하는 기관에 따라 다르다는 사실이다. 둘째는 우리나라의 여론조사 기관들이 조사방법론을 정확히 지키지 않는다는 사실을 지적하고 있다.

그러나 이러한 결과는 조사방법상의 차이로 인한 결과로 볼 수 있다. 실제로 여론조사에 사용하는 설문문항의 차이에 따라 응답자의 응답결과는 다르게 나타나게 된다. 또한 질문 형태에 따라서도 그 결과는 매우 달라지게 된다. 동일한 조사문항을 사용하는 경우라 할지라도 조사방식을 달리하는 경우 그 결과가 달라질 수 있다. 그러므로 이러한 오해를 제거하면서 정확한 여론조사의 결과를 제시하기 위해서는 보다 안정적이고 표준화된 여론조사의 방법이 절실히 요구된다고 하겠다.

현재 우리나라에서는 전화조사에 의한 여론조사방법을 주로 사용하여 왔으며 최근 들어 인터넷을 이용한 조사방법이 새롭게 대두되고 있다. 인터넷을 이용한 여론조사의 경우에는 전화조사에 비해 조사비용이 매우 저렴하고 또한 많은 양의 조사가 가능하다는 장점을 가지고 있다. 그러나 현재까지 인터넷사용자의 대부분이 20~30대의 젊은층이 주류를 이루고 있다는 단점을 가지고 있다. 즉 인터넷 조사의 추출틀은 조사 대상의 전체를 대표하지 못하고 일부 계층만을 포함하여 모집단의 구조적인 특징과 인터넷 조사에 이용되는 추출틀의 구조적 특성에 대한 연구 없이는 부분적인 추출틀의 사용으로 인한 비표본오차를 제어하기 힘들다. 때문에 전화조사의 추출틀의 구조적 특징들을 분석하여 이런 비표본오차의 문제점을 보완해 주는 연구가 필요하다.

사회여론조사나 선거여론조사 등에서 전화조사와 인터넷조사를 서로 결합하여 사용하는 방법은 서로 다른 추출틀을 이용한다는 관점에서 다중틀(Multiple Frame)을 이용한 조사라 할 수 있으며 또한 서로 다른 조사방법을 이용하므로 혼합조사법(Multi-Mode)이라 할 수 있다.

II. 다중 추출틀 조사기법

다중 추출틀을 활용한 조사기법을 사회여론조사나 선거여론조사 분야뿐만 아니라 보다 다양한 분야에서 활용하기 위해서는 먼저 주제를 이해하고, 이후에는 전화조사와 인터넷조사에서 사용할 각 추출틀에 대한 구조적인 특성을 파악하는 것이 필요하다. 예를 들어 사회인구학적인 특성인 성별-연령대별 구성비와 각 구성 요인별 특성을 파악하고 센서스자료나 기타 행정 통계자료를 조사대상 모집단의 구조적인 특성 파악의 보조자료로 활용하는 것이 바람직하다.

선정된 주제에 대한 지지율의 추정량과 표준오차의 계산식들을 유도하기 위한 기존의 연

구들을 살펴보자.

'Hartley(1974)'는 농업경제조사에서 농축산물의 생산량을 추정하기 위하여 농촌지역을 대상으로한 지역추출법과 농축산협회의 명부를 추출틀로 사용한 단순임의 추출법을 결합하여 최적가중절차(Optimum-Weight Procedure)를 사용하여 추정량을 다음과 같이 제안하였다.

$$\widehat{Y}_H = N_a \bar{y}_a + N_{ab}(p \bar{y}_{ab} + q \bar{y}_{ab}) + N_b \bar{y}_b$$

여기서 a, b, ab는 각 추출틀의 조사가구 단위를 나타내며 a, b는 중복되지 않은 부분을 그리고 ab는 추출틀이 중복된 부분의 조사단위를 나타낸다. 또한 $p+q=1$ 이다. 이때 추정량은 불편추정량이 된다. p 를 최적화하기 위하여 단순임의추출을 하였을 경우 사후총화추정량의 분산은 다음과 같이 주어졌다.

$$Var(\widehat{Y}_H) = (1-f_A) \frac{N_A^2}{n_A} \{ \sigma_a^2(1-\alpha) + p^2 \sigma_{ab}^2 \alpha \} + (1-f_B) \frac{N_B^2}{n_B} \{ \sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta \}.$$

여기서 $f_A = \frac{n_A}{N_A}$, $f_B = \frac{n_B}{N_B}$ 이며, $\alpha = \frac{N_{ab}}{N_A}$, $\beta = \frac{N_{ab}}{N_B}$ 이다.

총비용을 $C = C_A n_A + C_B n_B$ 와 같이 선형함수로 가정하면, 최적의 p , n_A , n_B 의 값은 각각 다음의 관계식을 통하여 얻을 수 있다.

$$\frac{C_A p^2}{C_B q^2} = \frac{(1-f_B)^2 (\sigma_a^2(1-\alpha) + p^2 \sigma_{ab}^2 \alpha)}{(1-f_A)^2 (\sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta)}$$

$$\frac{n_A}{N_A} = \left\{ \frac{\sigma_a^2(1-\alpha) + p^2 \sigma_{ab}^2 \alpha}{\lambda C_A} \right\}^{1/2}$$

$$\frac{n_B}{N_B} = \left\{ \frac{\sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta}{\lambda C_B} \right\}^{1/2}$$

여기서,

$$\lambda^{1/2} = \frac{1}{C} [N_A \{ \sigma_a^2(1-\alpha) + p^2 \sigma_{ab}^2 \alpha \} C_A]^{1/2} + N_B \{ \sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta \} C_B]^{1/2}$$

이다. 위의 결과로부터 'Lund(1968)'는 각 추출틀에서 단순임의추출법을 사용할 경우 p 의 최적값이 다음과 같음을 보였다.

$$p = \frac{\alpha n_A}{\alpha n_A + \beta n_B} = \frac{f_A}{f_A + f_B}$$

최근 Sharon과 Rao(2000)은 두 개의 추출틀조사의 경우에 대하여 모총계 Y에 대한 추정량의 이론적, 경험적인 성질에 대하여 조사하였으며 추정량에 대한 잭나이프(Jackknife) 분산추정량을 제안하고 이론적인 성질들에 관해 연구하였다. 또한 모의실험연구를 통해 분산에 대한 잭나이프 추정량이 매우 효과적임을 보였다.

그러나 전화조사와 인터넷조사를 이용한 사회 여론조사나 선거 여론조사의 경우에는 두 추

출처들이 서로 중복되는 부분이 없는 독립적인 경우로 가정하여 연구가 진행하여 앞의 연구와는 서로 상이한 상황이다.

본 연구에서는 이러한 서로 독립적인 출처들을 이용하는 경우를 가정한 전화조사와 인터넷조사의 출처들을 활용하여 결합추정량을 얻는 방법을 소개한다. 먼저 이들 두 출처들은 모두 성별-연령대별 구조에 따른 출처들을 이용하여 각각의 추정량을 산출하고, 이로부터 얻어진 각 추정량을 서로 결합하여 가중평균형태의 새로운 결합추정량을 얻는 방법을 제안한다. 결합추정량의 효율성을 조사하기 위하여 실제 자료를 통한 실증분석을 실시한다. 실증분석을 위해 지난 4월 13일 치뤄진 국회의원 선거에 대하여 한국갤럽조사연구소에서 실시한 전화조사와 인터넷조사자료를 이용하여 수행한다.

1. 조사방법별 추정량의 결합을 통한 결합추정법

전화조사에 대한 표본프레임과 인터넷조사의 표본프레임을 사용하는 경우, 이들 조사로부터 얻어진 각 추정값, \hat{X}_{tel} 과 \hat{X}_{int} 을 서로 결합한 결합추정량 \hat{X}_{com} 을 다음과 같이 일차결합의 형태로 표현할 수 있다.

$$\hat{X}_{com} = w_1 \hat{X}_{tel} + w_2 \hat{X}_{int} \quad <2.1>$$

여기서 w_1 과 w_2 는 각각의 방법을 통해 얻어진 추정값에 대한 가중값을 나타내며 다음의 관계식을 통하여 최적값을 얻을 수 있다. 즉 고정비용 C 에 대하여 분산 $Var(\hat{X}_{com})$ 를 최소화하는 w_1 과 w_2 값을 구함으로써 얻어질 수 있다.

$$\min_{w_1, w_2} Var(\hat{X}_{com}) = \min_{w_1, w_2} Var(w_1 \hat{X}_{tel} + w_2 \hat{X}_{int}) \quad <2.2>$$

여기서 고정비용 $C = E(C_{tel}) + E(C_{int})$ 로 전화조사의 평균비용과 인터넷조사의 평균비용을 더한 값이다. 그러므로 두 가중치의 값은 다음과 같이 전화조사와 인터넷조사에 의한 분산추정량의 비율로서 각각 얻어질 수 있다.

$$\hat{w}_1 = \frac{Var(\hat{X}_{int})}{Var(\hat{X}_{tel}) + Var(\hat{X}_{int})}$$

$$\hat{w}_2 = 1 - \hat{w}_1$$

2. 출처들의 성별-연령대 셀별 추정값의 결합을 통한 결합추정법

전화조사와 인터넷조사에 대한 각각의 추정치는 성별-연령대별에 따른 각 셀에 대한 추정치의 형태로 얻어진다. 그러므로 이들 추정치를 결합하는 또한가지 방법으로 이들 각 셀들에 대한 추정값을 서로 결합하여 합산함으로써 얻어질 수 있다. 즉

$$\hat{X}_{com} = \sum_{i=1}^k w_i \hat{X}_{comi}$$

여기서, 각 i 번째 셀별 결합추정량 \hat{X}_{comi} 는 다음과 같은 관계에 의해 얻어진다.

$$\hat{X}_{comi} = \alpha_{1i} \hat{X}_{teli} + \alpha_{2i} \hat{X}_{inti}$$

$$\alpha_{1i} = \frac{Var(\hat{X}_{inti})}{Var(\hat{X}_{teli}) + Var(\hat{X}_{inti})}$$

$$\alpha_{2i} = 1 - \alpha_{1i}$$

단, 이 방법의 경우 각셀의 빈도수가 5보다 작은 경우에는 각 셀의 분산추정량이 적절하지 않을 수 있으므로, 빈도가 적은 셀들을 서로 결합하여 주는 것이 바람직하다.

Ⅲ. 인터넷 조사와 전화조사 방법

1. 전화조사

전화조사는 면접원과 응답자가 전화를 이용하여 면접원이 설문문항을 제시하면 응답자가 그에 대한 응답을 하는 방식으로 이루어지는 조사방법의 형태이다. 그러므로 전화조사의 경우에는 전화의 보급률이 조사의 성패를 좌우하게 된다. 전화보급률이 100%에 미치지 못하는 경우에는 모집단을 구성하는 구성원이 무작위로 추출되지 않는다는 문제점을 가지게 된다. 우리나라의 경우에는 전화보급률에 있어서는 100%를 넘었으나 1가구에 1대 이상의 전화를 보유하고 있는 가구는 모집단 구성원에 중복이 되는 현상이 발생하므로 또한 문제점이 발생할 수 있다.

일반적으로 전화조사는 직접방문을 통한 개별면접의 경우와 비교해 볼 때 조사비용이 저렴하고 조사에 소요되는 시간을 단축할 수 있다는 장점을 지니고 있다. 그러나 전화를 이용해야 하므로 조사의 내용이 매우 제한적일 수밖에 없으며 실제 면접진행시 비대면적 상황이 응답의 내용을 왜곡시킬 수 있는 가능성이 높다는 단점을 가지고 있다. 또한 대체적으로 전화조사의 경우에는 무응답률이 높은 것으로 알려져 있다. 또한 부재자에 대한 적절한 교체표본을 선정하는 절차를 엄격히 준수한다는 것이 현실적으로 어렵다는 특징이 있다. 그러므로 전화조사의 결과에 대한 신뢰성과 보편성에 문제가 발생할 소지가 있을 수 있다.

전화조사의 모집단 리스트를 전화번호부를 이용하는 경우에는 다음과 같은 사항을 결정해야 한다. 먼저 전화번호부에서 어느 면을 리스트로 활용할 것인가의 문제를 결정해야 하며, 결정된 각 면에서 몇 명의 응답자를 추출할 것인가를 결정해야 한다. 또한 이 경우에는 최신의 전화번호가 공개되어 있지 않은 대상자의 경우에는 표본추출률에서 완전히 누락되는 결과가 발생하게 된다. 그러므로 인구이동률이 높은 신도시의 경우에는 많은 수의 대상자들이 전화번호부에서 누락되어 있을 수 있게 된다.

[표 1] 전화조사의 장·단점 비교

장 점	<ul style="list-style-type: none"> ● 개별면접에 비해 비교적 조사비용이 저렴하다 ● 조사에 소요되는 기간 단축이 가능하다 ● 제한된 지역을 대상으로한 조사에 적합하다
단 점	<ul style="list-style-type: none"> ● 모집단 구성 리스트를 얻기 어렵다. ● 조사내용이 매우 제한적으로 수용한다 ● 실제면접진행시 응답을 왜곡시킬 가능성이 높다 ● 무응답률이 높다 ● 부채시 대체표본을 선정하는 절차의 엄격한 준수가 어렵다

이상으로 볼 때, 전화조사의 경우에는 제한된 지역을 대상으로한 여론조사의 경우에는 유리하나 전국조사와 같은 광범위한 단위의 조사에는 부적절하다고 볼 수 있다. 또한 광범위 단위에 대한 조사의 경우에 주로 사용되는 전화조사방법은 random digit dialing와 같은 기법이 활용되고 있으며 이를 활용하는 경우에는 소요되는 시간과 인원 및 사전준비에 대한 절약이 가능하다는 장점을 가지고 있다 그러나 이 경우에도 통화성공률이 떨어진다는 단점이 있다.

2. 인터넷조사

최근 인터넷의 보급이 급속도로 확산되면서 이를 이용한 다양한 서비스의 제공이 활발하게 제공되고 있다. 또한 여론조사와 같은 사회조사의 경우에도 대량의 조사대상자를 손쉽게 얻을 수 있다는 장점으로 인해 인터넷을 활용한 조사방법이 활발히 개발되고 있는 실정이다. 정보통신부의 발표에 의하면 국내 인터넷 인구는 지난해 1,000만명을 넘어서고 있는 것으로 나타났다. 또한 일주일에 1회이상 인터넷을 이용하는 상시 이용자의 경우에도 7백 86만명에 이르는 것으로 조사되었다. 또한 인터넷 이용자의 66.3%는 1주일에 두세번 넘게 사용한다고 답해 인터넷의 이용빈도가 매우 높음을 나타내었다.

이러한 인터넷 이용자의 급진적인 증가와 더불어 이를 활용한 다양한 콘텐츠들이 새로이 개발되고 있는 실정이며, 여론조사의 경우에는 이들 인터넷 이용자를 효율적으로 활용할 수 있는 인터넷조사의 방식을 새로이 도입하기 시작하고 있다. 그러나 아직까지 인터넷 이용자의 대부분이 20~30대의 젊은 층에 편중되어 있는 경향이 있으며, 성별의 경우에도 남성이 여성보다 이용자수 면에서 훨씬 많은 비율을 차지하고 있어 아직까지는 조사목적에 적합한 모집단을 대표한다고는 볼 수 없는 실정이다. 또한 인터넷을 사용하는 지역이 대도시 중심이거나 학교나 관공서 그리고 일반기업 등과 같은 시설에 많이 치우친 경향을 보이고 있으며 인터넷 이용이 어려운 일명 '사이버 빈촌'이라는 지역을 대상으로 한 인터넷 조사가 현재까지는 불가능한 실정이다. 그러나 얼마전 정보통신부의 발표에 의하면 이러한 사이버 빈촌이나 중소도시를 대상으로 한 초고속 정보통신망의 보급실시로 인하여 앞으로 인터넷 사용인구는

급속히 증가할 것이며, 인터넷 이용자의 구성비 또한 연령이나 성별에 따른 차이를 보이지 않을 것이라 예상된다.

현재 인터넷조사를 활용한 여론조사는 국내의 몇몇 리치기관에 의해 이루어지고 있는 실정이다. 지난 4·13총선에서 한국갤럽은 실험적으로 선거여론조사를 인터넷을 이용하여 실시한 바 있다. 또한 인터넷 이용자를 대상으로 한 다양한 종류의 여론조사가 현재 각 리서치기관에 의해 수행되고 있다.

인터넷을 이용한 여론조사의 경우에는 기존의 개별면접조사나 전화조사와는 달리 응답자가 직접 해당 사이트에 접속하여 질문에 응답하는 형식을 이용하므로, 여론조사 기관에서는 응답자를 선정하기 위한 별도의 노력이나 비용을 들이지 않아도 되는 장점을 지니고 있다. 또한 짧은 시간동안에 많은 양의 자료를 수집할 수도 있으며 시·공간적인 제약을 받지 않는다는 장점을 가지고 있다. 이러한 인터넷조사의 장단점을 정리하면 [표 2]와 같다.

또한 인터넷조사의 경우 전자메일을 활용하여 선정된 대상자를 대상으로 한 조사도 가능하므로 매우 편리한 조사 방법이라 할 수 있다. 그러나 앞서 언급한 바와 같이 현재까지 인터넷 사용자의 구성비가 균일하지 않다. 즉 연령면으로 볼 때 20~30대 젊은 층이 주류를 이루고 있으며, 성별에 있어서도 남자의 구성비가 여자에 비해 높은 편이라 정확한 조사결과를 얻기가 어렵다는 단점을 지니고 있다. 지난 4·13총선의 결과에서도 인터넷 이용자를 대상으로 한 조사에서는 일반적으로 민주당에 대한 지지율이 한나라당의 지지율보다 높은 경향을 보이는 것으로 조사되었으나 결과는 정반대로 나오는 예가 있었다. 이는 인터넷을 사용하는 층이 주로 20~30대의 젊은 층인 것이 원인이었다. 또한 인터넷 이용자가 아닌 대상자에 대한 정보를 얻을 수 없다는 단점도 가지고 있다.

[표 2] 인터넷조사의 장·단점 비교

장 점	<ul style="list-style-type: none"> ● 개별면접과 전화조사에 비해 비용이 적게 든다 ● 조사에 소요되는 기간이 짧다 ● 많은 양의 자료를 얻을 수 있다 ● 시·공간적인 제약을 받지 않는다 ● 응답자와의 대화형식을 통한 자료수집이 가능하다. ● 응답자의 반응에 대한 상호대응이 자유롭다. ● 그래픽·음성·동영상을 활용한 질문지의 작성이 가능하다.
단 점	<ul style="list-style-type: none"> ● 인터넷 이용자가 젊은층으로 치우쳐 있다 ● 인터넷 이용자의 성별 구성비가 다르다 ● 사이트 방문자에 의해 이루어지므로 대상자가 고르지 못하다. ● 인터넷 미사용자에 대한 자료수집이 불가능하다

IV. 사례조사 자료분석

지난 2000년 4월 13일 실시된 국회의원총선거가 실시되었다. 다음은 4·13총선에서 한국갤럽조사연구소에서 실험적으로 실시한 전화여론조사결과와 인터넷여론조사의 결과이다. 이들 결과를 바탕으로 전화조사와 인터넷조사의 결과를 비교하여 보기로 한다.

본 연구에서는 전체 조사대상지역 중에서 서울의 강남갑, 관악갑, 성동구, 송파갑, 그리고 전주완산 선거구 등의 5개 선거구를 선정하였다. [표 3]의 결과는 5개 선거구에 대한 인터넷 조사와 전화조사를 위해 사용된 표본수를 나타낸 것이다. 추출된 표본을 살펴보면 전화조사와 인터넷조사의 일반적인 경향을 잘 반영하고 있음을 알 수 있다. 즉 인터넷 조사의 경우에는 젊은층인 20~30대의 연령층의 표본이 각 선거구의 표본 중의 약 60%가량이 추출되어 상당히 많은 구성비를 이루고 있으며, 전화조사의 경우에는 30~50대 이상의 표본수에 고루 분포되어 있는 것을 볼 수 있다.

[표 3] 전화조사와 인터넷조사에 사용된 각 선거구별 표본수 및 구성비

구 분	강남갑		관악갑		성동구		송파갑		전주완산			
	표본수 (명)	구성비 (%)	표본수 (명)	구성비 (%)	표본수 (명)	구성비 (%)	표본수 (명)	구성비 (%)	표본수 (명)	구성비 (%)		
인터넷 조사	남	20대	108	41	74	35	90	39	79	34	86	39
		30대	46	18	50	24	39	17	41	18	35	16
		40대	16	6	7	3	12	5	16	7	26	12
		50대이상	13	5	9	4	1	0	15	6	10	5
	여	20대	56	21	59	28	78	34	67	29	51	23
		30대	19	7	9	4	9	4	8	3	10	5
		40대	3	1	2	1	1	0	4	2	2	1
		50대이상	1	0	2	1	1	0	2	1	1	0
소 계		262	100	212	100	231	100	232	100	221	100	
전화 조사	남	20대	61	13	35	8	44	8	41	9	48	12
		30대	42	9	32	8	65	12	30	7	44	11
		40대	40	9	54	13	64	12	46	10	34	8
		50대이상	101	22	88	21	103	19	95	22	81	20
	여	20대	35	8	44	11	39	7	31	7	56	14
		30대	49	11	56	13	79	14	57	13	56	14
		40대	65	14	40	10	60	11	71	16	43	10
		50대이상	73	16	66	16	97	18	70	16	51	12
소 계		466	100	415	100	551	100	441	100	413	100	

[표 4] 선거구별 표본수에 대한 가중치

구 분		강남갑		관악갑		성동구		송파갑		전주완산			
		가중치 (명)	구성비 (%)	가중치 (명)	구성비 (%)	가중치 (명)	구성비 (%)	가중치 (명)	구성비 (%)	가중치 (명)	구성비 (%)		
인터넷조사	남	20대	41	15.6	35	16.4	33	14.3	31	13.4	30	13.6	
		30대	28	10.6	28	13.1	34	14.7	26	11.3	28	12.7	
		40대	23	8.7	18	8.5	22	9.5	26	11.3	23	10.4	
		50대이상	33	12.5	26	12.2	27	11.7	27	11.7	25	11.3	
	여	20대	44	16.7	33	15.5	32	13.9	34	14.7	29	13.1	
		30대	28	10.6	24	11.3	29	12.6	28	12.1	29	13.1	
		40대	30	11.4	20	9.4	21	9.1	30	13.0	24	10.9	
		50대이상	36	13.7	29	13.6	33	14.3	29	12.6	33	14.9	
	소 계		263	100.0	213	100.0	231	100.0	231	100.0	221	100.0	
	전화조사	남	20대	74	15.9	68	16.4	79	14.3	60	13.6	56	13.6
			30대	50	10.7	54	13.0	80	14.5	50	11.3	52	12.6
			40대	41	8.8	36	8.7	52	9.4	50	11.3	43	10.4
50대이상			58	12.4	50	12.0	65	11.8	52	11.8	47	11.4	
여		20대	78	16.7	65	15.7	76	13.8	64	14.5	55	13.3	
		30대	49	10.5	47	11.3	70	12.7	52	11.8	55	13.3	
		40대	53	11.4	38	9.2	51	9.3	58	13.2	44	10.7	
		50대이상	63	13.5	57	13.7	78	14.2	55	12.5	61	14.8	
소 계		466	100.0	415	100.0	551	100.0	441	100.0	413	100.0		

그러므로 인터넷조사에 의한 후보지지율은 젊은층에 대한 의견이 집중적으로 반영되는 반면, 전화조사에 의한 결과는 전체 연령층의 의견이 고루 반영된 결과로 추측할 수 있을 것이다. 또한 다음의 [표 4]는 5개 선거구별 성별-연령별 가중치값이다. 이 가중치는 앞의 [표 3]의 선거구별 추출표본수를 기반으로 하여, 각 선거구별 성별-연령대별 유권자의 구성비로 하여 계산한 결과이다. 이를 [표 3]의 경우와 비교하여 볼 때, 인터넷 조사의 경우에는 가중치와 표본수의 구성비는 매우 상이한 분포형태를 보이고 있는 반면, 전화조사의 경우에는 가중치와 표본수의 구성비가 비교적 유사한 분포형태를 보이고 있는 것을 볼 수 있다.

1. 선거구별 전화조사와 인터넷조사 결과

다음의 [표 5]~[표 9]는 4.13 국회의원 선거에 즈음하여 조사된 한국갤럽조사연구소에서 발표한 서울의 강남갑, 관악갑, 성동구, 송파구선거구 그리고 전주완산 선거구에 대한 전화조사와 인터넷조사를 통해 얻어진 각 선거구의 후보자별 지지자수와 지지율을 얻은 결과표이

다. 계산의 편의를 위하여 각 선거구에 등록된 후보들 중 전화조사와 인터넷조사결과에서 비교적 높은 지지율을 보인 4명 또는 5명의 후보만을 고려하여 추정을 실시하였다.

[표 5] 강남갑 선거구의 인터넷 조사자료와 지지율(262명)

구 분		한나라당 최 병 렬	민 주 당 전 성 철	자 민 련 김 명 년	진 보 당 박 윤 기
남	20대	41	42	1	6
	30대	11	29	0	0
	40대	8	7	0	0
	50대이상	6	6	0	0
여	20대	18	21	1	6
	30대	10	5	0	0
	40대	2	1	0	0
	50대이상	1	0	0	0
합 계		97	111	2	12
인터넷조사 지지율		42.4	48.5	0.9	5.3
조정 지지율		55.8	38.7	0.5	2.9
전화조사 지지율		53.8	31.2	6.2	5.1
개표결과(4/13)		56.5	36.8	3.1	1.9

[표 6] 관악갑 선거구의 인터넷 조사자료와 지지율(212명)

구 분		한나라당 김 성 식	민 주 당 이 훈 평	자 민 련 이 상 현	진 보 당 박 용 덕
남	20대	15	31	8	6
	30대	13	18	4	5
	40대	1	3	0	0
	50대이상	4	3	2	0
여	20대	7	17	5	7
	30대	2	4	0	0
	40대	0	1	1	0
	50대이상	0	1	0	0
합 계		42	78	20	18
인터넷조사 지지율		25.2	46.7	12.0	10.8
조정 지지율		22.7	51.5	14.8	5.7
전화조사 지지율		26.3	36.9	22.5	3.0
개표결과(4/13)		29.9	43.9	17.1	3.1

[표 7] 성동구 선거구의 인터넷 조사자료와 지지율(231명)

구 분		한나라당 이 세 기	민 주 당 임 종 석	자 민 련 안 승 근	진 보 당 최 대 업	무 소 속 김 형 곤
남	20대	16	50	1	4	14
	30대	6	25	0	3	1
	40대	6	4	1	1	0
	50대이상	0	1	0	0	0
여	20대	11	45	0	4	8
	30대	4	2	0	0	1
	40대	0	1	0	0	0
	50대이상	0	0	0	0	0
합 계		43	128	2	12	24
인터넷조사 지지율		20.4	60.7	0.9	5.7	11.4
조정 지지율		21.6	65.8	1.2	4.1	6.9
전화조사 지지율		37.6	40.4	6.5	3.9	10.3
개표결과(4/13)		38.9	48.4	2.3	1.4	8.7

[표 5]에서 [표 9]까지의 각 지역구별 전화조사와 인터넷조사의 결과를 살펴보면 인터넷조사의 경우에는 민주당후보의 지지율이 개표결과의 지지율보다 매우 높게 추정되어진 결과를

[표 8] 송파갑 선거구의 인터넷 조사자료와 지지율(232명)

구 분		한나라당 맹 형 규	민 주 당 김 영 술	자 민 련 전 의 성	진 보 당 강 인 성
남	20대	28	33	1	2
	30대	12	20	0	4
	40대	9	4	0	0
	50대이상	8	7	0	0
여	20대	21	17	0	8
	30대	3	4	0	0
	40대	2	1	0	0
	50대이상	1	1	0	0
합 계		84	87	1	14
인터넷조사 지지율		43.9	45.6	0.5	7.3
조정 지지율		49.4	44.9	0.2	3.7
전화조사 지지율		50.1	35.9	7.5	2.8
개표결과(4/13)		54.8	37.2	3.7	2.3

보이는 반면 전화조사의 경우에는 그렇지 않은 것을 보여주고 있다. 이는 인터넷 사용자의 연령층이 20~30대의 젊은층이 주류를 이루므로 이러한 성향이 인터넷 조사결과에 반영된 것이라 하겠다. 또한 전화조사의 경우에는 실제 개표결과에는 어느 정도 근사하나 다소 과소 적합되어지는 경향이 존재하는 것을 볼 수 있다. 그러므로 이들 두 가지 조사방법에 의한 결과를 두 개의 독립적인 추출들로 간주하여 서로 결합하는 방법을 사용하여 보완하는 방법을 사용하는 것은 매우 효과적이라 할 수 있다.

[표 9] 전주완산선거구의 인터넷 조사자료와 지지율(221명)

구분		한나라당 임 광 순	민 주 당 장 영 달	무 소 속 김 회 진	무 소 속 장 세 환
남	20대	0	42	15	8
	30대	1	18	6	6
	40대	0	12	6	5
	50대이상	1	3	2	3
여	20대	0	18	11	6
	30대	0	2	4	0
	40대	0	2	0	0
	50대이상	0	1	0	0
합 계		2	98	44	28
인터넷조사 지지율		1.2	57.0	25.6	16.3
조정 지지율		1.8	65.1	20.7	12.4
전화조사 지지율		4.7	54.8	28.5	12.0
개표결과(4/13)		5.0	66.7	18.6	9.8

전화조사와 인터넷조사의 두가지 추출들의 추정량을 서로 결합한 결합추정량은 두 추출들에서 얻어진 추정량에 대한 가중평균의 형태를 가지게 되는데, 이때 각 가중치를 결정하는 방법을 다음과 같은 두가지 방식에 의하여 접근하도록 한다.

2. 실증자료의 추정결과

다음의 결과는 지난 4.13 총선거 기간동안 한국갤럽(주)에서 실시한 전화조사와 인터넷조사의 결과를 바탕으로 하여 앞서 제안한 두가지 결합방법을 각각 적용하여 얻어진 결과이다. 전체 지역구들 중에서 서울의 강남갑, 관악갑, 성동갑, 송파갑 선거구와 전주완산선거구에 대한 추정결과이다.

[표 10] 강남갑선거구의 결합추정결과

구분	한나라당 최병렬	민주당 전성철
결합추정1	154	90
결합추정2	155	91
추정지지율1	58.4	34.2
추정량분산1	0.00061	0.00056
추정지지율2	57.5	33.5
추정량분산2	0.00060	0.00055
인터넷조사	55.8	38.7
전화조사	61.0	28.8
개표결과	56.5	36.8

[표 11] 관악갑선거구의 결합추정결과

구분	한나라당 김성식	민주당 이훈평
결합추정1	44	94
결합추정2	47	97
추정지지율1	23.3	49.7
추정량분산1	0.00066	0.00092
추정지지율2	23.2	47.7
추정량분산2	0.00051	0.00071
인터넷조사	22.7	51.5
전화조사	24.3	46.9
개표결과	29.9	43.9

[표 12] 성동구선거구의 결합추정결과

구분	한나라당 이세기	민주당 임종석
결합추정1	65	131
결합추정2	94	130
추정지지율1	29.0	58.2
추정량분산1	0.00065	0.00077
추정지지율2	37.2	51.3
추정량분산2	0.00054	0.00058
인터넷조사	21.6	65.8
전화조사	42.5	47.4
개표결과	38.9	48.4

[표 13] 송파갑선거구의 결합추정결과

구분	한나라당 맹형규	민주당 김영술
결합추정1	133	88
결합추정2	161	90
추정지지율1	55.3	36.5
추정량분산1	0.00082	0.00077
추정지지율2	59.8	33.3
추정량분산2	0.00061	0.00057
인터넷조사	49.4	44.9
전화조사	65.8	27.4
개표결과	54.8	37.2

[표 14] 전주완산선거구의 결합추정결과

구분	한나라당 임광순	민주당 장영달	무소속 속진
결합추정1	5	133	46
결합추정2	6	132	49
추정지지율1	2.5	63.6	22.2
추정량분산1	0.00009	0.00082	0.00061
추정지지율2	3.1	62.3	23.0
추정량분산2	0.00009	0.00069	0.00052
인터넷조사	1.8	65.1	20.7
전화조사	3.6	60.5	24.7
개표결과	5.0	66.7	18.6

[표 10] ~ [표 14]의 결합추정결과를 살펴보면, 5개 선거구에 대하여 모두 인터넷조사와 전화조사의 추정결과를 서로 절충하여 개표결과에 보다 근접한 추정결과를 얻어짐을 알 수 있다. 또한 두 결합추정값에 대한 분산추정량이 작고, 두 결합추정값과도 거의 차이를 보이고 있지 않으므로 인터넷조사와 전화조사의 결과를 이용하여 새로운 추정량을 얻는 결합추정1과 결합추정2의 방법이 모두 효과적이라는 사실을 보여주고 있다.

V. 결 론

현재 여론조사에 주로 이용되고 있는 전화조사 방법은 인터넷조사에 비하여 조사기간이 오래 걸리고 비용 또한 많이 들어 효과적이지 못하다는 단점을 가지고 있다. 그러나 인터넷 조사와는 달리 표본 프레임의 구성이 용이하다. 그러므로 이러한 장점을 활용하고 반면에 단점을 극복할 수 있는 방법은 비교적 고비용이며 조사기간이 긴 전화조사 방법을 1회 실시하고, 저비용과 단기간의 조사를 필요로 하는 인터넷 조사를 수회에 걸쳐 실시하여 이들 결과를 서로 결합하여 정확한 추정량을 얻는 결합추정방법의 연구가 적절하다고 할 수 있다. 이러한 결합추정방법으로써 본 연구에서는 인터넷조사에 의한 추정량과 전화조사에 의한 추정량을 서로 결합하여 새로운 추정량을 얻는 결합추정법과 표본프레임의 각 셀별로 인터넷 조사에 의한 추정량과 전화조사에 의한 추정량을 서로 결합한 다음, 이들 결과를 합하여 전체 추정량을 얻는 결합추정방법을 각각 제안하였으며, 지난 4.13 총선거에서 한국갤럽조사연구소가 실시한 전화조사와 인터넷조사자료를 바탕으로 이들 두 가지 방법을 적용하여 추정량을 얻은 결과 보다 개선된 추정량을 얻을 수 있음을 사례연구를 통하여 두가지 결합추정법이 모두 바람직한 결과를 도출할 수 있음을 보였다.

〈참 고 문 헌〉

1. 김영원, 변종석(2000), “인터넷조사에서 표본추출동향 및 문제점”, 《한국조사통계연구회 자료집》
2. 이계오(2000), “인터넷 여론조사의 현황과 전망”, 《한국조사통계연구회자료집》
3. Hartley, H. O.,(1974), “Multiple Frame Methodology and Selected Applications, Sankhya”:The Indian Journal of Statistics, Vol. 36, Series C. Pt. 3, pp. 99-118.
4. Rund, R. E. (1968), “Estimators In Multiple Frame Surveys”, Proceedings of the Social Statistics Section, American Statistical Association, 282-288.
5. Sharon L. Lohr and Rao, N. K.,(2000), “Inference From Dual Frame Surveys, Journal of the American Statistical Association”, Vol. 95, No. 449, pp. 271-279.
6. Skinner C. J. and Rao, J. N.,(1996), “Estimation in Dual Frame Surveys with Complex Design”, Journal of the American Statistical Association, Vol. 91, No. 433, pp. 349-356.
7. Suchada Seidhoranajul Kiranandana, (1976), “Imperfect Frames In Sample Surveys”, Ph. D thesis, Harvard University.