

연구논문

2012년 총선 출구조사 예측오차 분석*

Analysis of Forecasting Error of the Exit Poll for the General Election of 2012 in Korea

곽은선^{a)} · 김지연^{b)} · 김영원^{c)}

Eun-Sun Kwak · Ji-Yeon Kim · Young-Won Kim

본 연구에서는 출구조사 예측오차의 발생 원인과 특성을 알아보기 위해 2012년에 치른 19대 총선 출구조사의 자료를 이용하여 출구조사 예측오차를 분석하였다. 예측 오차를 정리해 본 결과 새누리당에 대해서는 과소 예측이, 통합야당에 대해서는 과대 예측이 발생했으며, 시·도별로 편향 수준에 차이가 있었다. 예측오차를 투표소추출 오차와 실사오차로 구분해 보면 투표소추출오차에 비해 실사오차가 더 심각하였다. 또한 오차를 발생시킬 것으로 예상되는 변인들과 투표소추출오차 및 실사오차와의 상관분석을 통해 조사거절률이 실사오차에 유의한 양의 상관관계를 갖고 있으며, 표본 추출률을 증가시키면 투표소추출오차가 감소한다는 사실도 확인할 수 있었다. 한편 출구조사를 진행한 세 개 조사기관별로 오차를 비교해 본 결과 특히 실사오차에 있어서 조사기관간의 차이가 통계적으로 유의한 것으로 나타났다.

주제어: 출구조사 예측오차, 투표소추출오차, 실사오차, 조사거절률.

In this study, we investigate the data of the exit poll for the 19th general election of 2012 to analyze the causes and characteristics of exit poll forecasting error. The empirical studies show that systematic negative bias of candidates affiliated to Saenuri Party is serious in the specific regions. To explore the main causes and effects of the forecasting error, the total survey error is divided by

* 본 연구는 숙명여자대학교 2011학년도 교내연구비 지원에 의해 수행되었음.

a) 숙명여자대학교 통계학과 박사과정

b) 밀워드브라운미디어리서치 상무이사

c) 교신저자(corresponding author): 숙명여자대학교 통계학과 교수 김영원.

E-mail: ywkim@sm.ac.kr

sampling error of the polling sites and non-sampling error which comes from selecting voter and collecting responses from sampled voters in each polling sites. And we consider relationships among several factors, which are anticipated as main causes of error, and non-sampling error as well as sampling error. As a result of correlation analysis, we can figure out that the non-sampling error of the exit poll mainly come from non-response. Also we consider the house effect in exit poll.

Key words: exit poll, forecasting error, sampling error, non-sampling error

I. 서론

선거예측조사는 선거 전에 전화를 이용한 조사를 통해 후보자의 득표율을 예측하는 전화여론조사와 투표 당일 투표소에서 투표를 마치고 나오는 사람들을 조사해 선거결과를 예측하는 출구조사로 구분할 수 있다. 본 연구에서는 2012년에 치른 19대 총선의 출구조사 자료를 이용해 해당 출구조사에서 발생한 오차의 특징과 함께 그 오차의 발생 요인에 대해 분석해 보고자 한다.

통계학적 관점에서는 선거예측조사에서 발생한 오차가 어느 정도 수준인지와 함께 오차를 발생시킨 요인이 무엇인지, 그리고 그 결과를 바탕으로 향후 선거예측조사에서 보다 정확한 예측을 하기 위해서는 어떤 점을 개선해야 할 것인지에 대해 주로 관심을 갖게 된다. 여기서 선거예측조사의 정확성은 특정 후보의 지지율에 대한 예측값과 참값(개표결과)의 차이로 설명할 수 있다. 출구조사의 경우 투표 당일의 투표자를 대상으로 조사가 이루어지며, 투표가 끝나는 6시에 방송을 통해 예측 결과를 발표하게 된다. 이 때 각 후보자에 대한 예측 득표율과 개표가 끝난 후 집계되는 최종 득표율과의 비교를 통해 출구조사의 정확성을 평가하게 된다.

정치권에서는 각 정당이 몇 개의 의석을 차지하게 될 것인지에 대해서, 또는 후보자 개인 입장에서는 당락 여부 등이 관심대상이 될 수 있지만, 이준웅(2004)이 지적한 것처럼 당선자를 잘못 예측한 선거구 수를 기준으로 선거예측조사의 정확성을 단순히 평가하는 것은 과학적인 접근으로 볼 수 없다. 실제 출구조사는 상당히 복잡한 조사

및 분석 과정을 통해 선거예측 결과를 산출하게 된다. 따라서 출구조사의 전반적인 수행과정을 살펴보면 투표소 추출방법뿐만 아니라, 조사원 관리 및 현장실사 방법, 조사 거절과 거짓응답, 표본투표소에서의 응답자 선정방법, 무응답 및 가중치 처리 등 다양한 원인에 의해 예측오차가 발생함을 알 수 있다. 따라서 출구조사의 정확성 제고를 위해서는 출구조사에서 발생하는 오차의 특성을 정확히 이해하고 이를 기반으로 현실적인 한계를 감안한 개선방안이 모색될 필요가 있다(김영원·곽은선 2010).

2012년 19대 총선 이전까지의 선거예측조사에서는 부족한 예산과 조사회사의 실사 운영 능력의 한계 등을 고려하여, 사전에 전화조사를 통해 각 지역구별 경합도를 분석하고, 경합도가 높은 지역에 대해서는 출구조사를 수행하되 나머지 지역에 대해서는 전화조사를 하는 방식으로 이루어져 왔다. 실제 이런 방식의 총선 출구조사를 수행한 결과, 경합도 분석에 실패하여 경합도가 높은 지역을 전화조사 대상으로 분류하는 오류와 함께 특히 전화조사를 실시한 지역구에서 상당히 큰 오차가 발생하였다. 따라서 대선이나 지방선거와 달리 총선의 경우 성공적인 선거예측조사 결과를 얻지 못했던 경험을 갖고 있다.

이런 문제점을 고려하여 2012년 19대 총선의 경우 모든 246개 선거구에서 출구조사를 수행하였다. 이와 같이 전 지역 출구조사가 가능했던 것은 한국방송협회 주관으로 지상파 방송 3사가 공동예측조사위원회(Korea Election Pool; 이하 KEP)를 구성하여 공동 조사를 실시함으로써 전 지역 출구조사에 필요한 예산 및 실사운영 능력을 확보할 수 있었다는 점과 방송 3사가 공동 출구조사를 통해 2010년 지방선거에서 정확하게 예측함으로써 세간을 놀라게 했던 경험 등이 원동력이 되었다. 특히 대선이나 지방선거와는 달리 그동안 총선 선거예측조사에서 경험했던 개별 선거구별 예측과 함께 전체 정당별 의석수 예측에 있어서 발생하는 어려움 등이 전 지역 출구조사를 수행하게 된 중요한 요인이었다.

2012년 총선 선거예측조사를 위해, 2010년 지방선거와 마찬가지로 방송협회 주관으로 KEP를 구성하여 전반적인 조사 설계뿐만 아니라 개별 지역구 후보자별 득표율 예측까지 전 과정을 방송3사 공동으로 실시하였다. 다만 정당별 의석수는 지역구별 득표율을 기초로 각 방송사가 별도의 추가적인 통계분석을 통해 방송사별로 결과를 예측하였다. 전체 선거예측 과정은 그동안 우리나라에서 실시된 대부분의 방송사 출구조사를 담당했던 밀워드브라운미디어리서치(이하 MBMR), 코리아리서치(이하 KRC), 그리고 TNS 3개 회사에서 수행하였다. 총선 선거예측을 위해 전국 246개 선거구에서 추

출된 투표소 수는 2,492개이고, 그동안의 출구조사 과정을 통해 얻었던 경험과 정립된 방법론을 적용하여 과거 선거예측조사보다도 예측의 정확성을 높이려고 최대한 노력하였다.

하지만 선거예측결과는 예상과는 달리 만족스럽지는 않았다. 특히 3개 방송사의 주요정당에 대한 의석수 예측결과는 실제 개표결과를 많이 빗나갔다. 2012년 총선 출구조사 예측결과와 실제 개표결과를 비교했을 때 확연히 들어난 현상은 새누리당에 대해서는 과소 예측을, 민주통합당에 대해서는 과대 예측을 하였다는 것이다. 특히 특정 정당 후보자의 득표율을 체계적으로 과대 또는 과소 추정함으로써 정당별 의석수 예측에 있어서는 큰 어려움이 발생하였고, 결과적으로 이번 선거에서도 의석수 예측에 있어서 원하는 수준의 정확한 결과를 얻지 못하였다.

많은 노력과 비용이 투자되었음에도 주요 정당의 의석수를 정확히 예측하지 못한 것은 방송사와 조사회사 모두에게 만족할 수 없는 결과였다. 그러나 모든 선거구에서 표본투표소를 추출하여 얻은 19대 총선 출구조사 자료는 앞으로 출구조사의 정확성 제고를 위한 연구에 중요한 자료가 될 것임이 분명하다.

그동안 우리나라의 대통령선거, 국회의원선거 및 지방선거에서의 출구조사에 있었던 오차의 발생원인을 분석하기 위해 수행된 연구결과들을 살펴보면, 허명희·홍내리(2000)는 당시 방송사 출구조사가 갖고 있는 전반적인 문제점을 파악하기 위해 적정 표본투표소 수와 투표소 표본추출 방법, 조사거절과 거짓응답 등에 따른 오차발생 수준과 출구조사 방법의 개선 필요성을 통계적 모의실험을 통해 보여 주었다. 기존 출구조사 투표소 추출방법의 개선 필요성과 관련해서는 조성검·김지연(2004)이 기존에 사용되던 ‘대표구추출법’ 대신 확률표본추출법의 도입이 필요하다는 연구결과를 제시하였으며, 이와 관련해 김영원·엄윤희(2005)은 17대 총선 개표자료를 토대로 효율적인 투표소 표본추출을 위한 정렬계통추출법을 제안하고 그 활용 가능성 및 효율성을 분석하였다.

출구조사의 오차발생 원인을 규명하기 위해 김영원·김지현(2007)은 2006년 지방선거 출구조사 자료를 기초로 투표자의 특성에 따른 투표성향과 시간대별 투표자 특성을 정리하고 출구조사의 예측오차를 연구하는 동시에 편향발생 원인 중 하나가 될 수 있는 최종 투표자 표본의 대표성을 검토하였다. 김영원·곽은선(2010)은 18대 총선의 출구조사 자료를 이용하여 총조사오차의 관점에서 출구조사의 정확성을 평가하였으며, 김영원·최윤정(2011)은 2010년 지방선거, 2006년 지방선거, 2007년 대통령선거의 출

구조사 자료를 토대로 출구조사 응답과정에서 시·도 단위에서 특징적으로 발생하는 지역별 편향에 대한 분석결과를 제시하고 있다. 박진우·배형민(2012)은 2010년의 지방선거에서 각 투표소에서 투표자 선정을 위해 적용한 추출간격에 따른 표본크기의 적절성에 대한 연구결과를 제시하였다. 한편 총선 출구조사의 정당별 의석수 예측과 관련해 허명회(2008)는 2004년 제17대 국회의원 선거자료를 이용하여 베イズ 이론(Bayes theory)을 기반으로 한 정당별 의석수 예측방법을 제시하고 관련 이론을 정리하였다.

본 연구에서는 2012년 4월 11일 진행된 제19대 총선 방송사 공동출구조사 자료를 이용하여 출구조사 선거구별 예측오차를 투표소추출오차와 실사오차로 분리하고, 각 오차가 16개 시·도별로 어떤 경향을 보이고 있는지 살펴보고 이들 오차의 발생 원인을 탐색하기로 한다. 출구조사의 정확도를 높이기 위해서는 오차의 발생 원인을 알아내고, 이러한 요인을 효과적으로 관리하는 것이 무엇보다 중요하다. 오차의 발생 원인을 분석하기 위해 출구조사 예측오차에 영향을 줄 것으로 예상되는 요인을 선정하여 이들 변인과 오차와의 연관성을 분석하였다.

한편 선거예측오차를 검토할 때, 적용된 조사방법론과 함께 조사기관에 따른 영향(house effect)도 고려할 필요가 있다. 조사기관의 고유한 조사방법 및 분석방법의 차이 때문에 예측치가 달라질 수 있기 때문이다. 따라서 19대 총선 출구조사를 담당한 세 개 조사기관(MBMR, KRC, TNS)간에 예측오차, 투표소추출오차 및 실사오차가 어떤 차이가 있는지를 알아보기 위한 통계분석도 별도로 수행하였다.

II. 출구조사 정당별 예측오차

선거예측조사에서 예측오차는 특정 후보의 예측 득표율($\hat{p}$)과 실제 득표율(p)의 차이를 의미하며, 이런 예측오차는 다음과 같이 표현할 수 있다.

$$\text{예측오차} = \hat{p} - p$$

예측 득표율이 실제 득표율에 비해 클 때 과대 예측이라 하고, 그 반대를 과소 예측이라 한다. <표 1>은 전국 246개 선거구 중 새누리당 후보가 1위나 2위를 한 210개 선거구와 민주통합당 또는 통합진보당(이하 통합야당) 후보가 1위나 2위를 한 205개 선

거구를 분석대상으로 하여, 각 정당별로 시·도 내의 평균 선거구별 예측오차를 계산해 정리한 것이다. 새누리당의 경우 210개 선거구를 분석해 보면 16개 시·도 중 9개 지역(서울, 부산, 대구, 인천, 대전, 울산, 경기, 강원, 충북)에서 새누리당 후보자의 득표율을 과소하게 예측한 것으로 나타났다. 예를 들어, <표 1>에서 서울의 경우 새누리당이 1위나 2위를 한 선거구는 46개이고 이들 선거구별 예측오차를 평균하면 -1.81%p로 나타나 서울에서는 새누리당 후보가 과소 추정되었다는 것을 알 수 있다. 한편, 통합야당이 1위 혹은 2위를 한 선거구는 전국 246개 선거구 중 205개 선거구였으며, 이들 선거구를 분석해 보면 16개 시·도 중 광주와 충남을 제외한 14개 시·도에서 통합야당 후보의 득표율이 과다하게 예측된 것으로 나타났다. 결국 19대 총선 출구조

<표 1> 16개 시·도별 정당별 예측오차 현황

(단위: 개, %)

시·도	새누리당		통합야당	
	선거구 수	예측오차	선거구 수	예측오차
서울	46	-1.81	48	2.20
부산	18	-0.56	17	1.37
대구	12	-1.81	7	1.86
인천	12	-0.86	12	1.57
광주	1	0.60	4	-1.36
대전	5	-0.68	5	1.63
울산	6	-0.93	6	1.99
경기	52	-0.68	51	1.46
강원	9	-0.79	8	1.77
충북	8	-0.30	8	0.83
충남	8	0.63	6	-1.45
전북	2	1.04	9	1.93
전남	0	-	7	1.95
경북	15	0.84	5	2.03
경남	15	0.62	9	1.19
제주	1	0.11	3	1.78

사에서 광주와 충남을 제외한 14개 시·도에서 새누리당에 대해서는 과소 예측하는 경향이, 통합야당에 대해서는 과대 예측하는 경향이 있었음을 확인할 수 있었다. 이런 현상과 관련해 김영원·최윤정(2011)은 지방선거나 대통령선거 출구조사에서도 시·도 단위에서 지역의 정치적 특성에 따라 지속적으로 여당과 야당 후보에 대한 과대 혹은 과소 추정 결과가 발생하고 있다는 점을 지적한 바 있다.

Ⅲ. 출구조사에서 투표소추출오차와 실사오차

출구조사는 전체 투표소 중에서 일부 투표소를 추출하여 선거 당일에 해당 투표소에서 투표를 마치고 나온 투표자를 대상으로 진행된다. 이때 해당 투표소에서 투표를 마치고 나온 모든 투표자가 출구조사의 대상이 되는 것은 아니며 계통표본추출 방식에 의해 일정 간격을 두고 조사하게 된다(19대 총선의 경우 5명 중 1명을 계통추출하였음). 그러나 조사자로 선정된 투표자가 모두 성실히 조사에 임하는 것은 아니며, 조사 자체를 거절하는 경우와 거짓응답을 하는 경우가 발생할 수 있다. 또한 5시까지의 조사결과를 토대로 선거결과를 예측해야 한다는 측면과 부재자 투표자는 출구조사 대상에서 제외된다는 제약점도 갖고 있다. 아울러 조사를 담당하는 조사원들이 조사회사로부터 사전에 교육받은 실사지침을 잘 이행하지 않을 수도 있다. 이와 같이 출구조사에서는 매우 다양한 원인에 의해 오차가 발생할 수 있다. 결국 출구조사 오차는 (1) 전체 투표소 중 일부 투표소를 추출하면서 발생하는 오차, (2) 추출된 투표소에서 일부 투표자를 선택하는 과정에서 발생하는 오차, (3) 그밖에 조사대상으로 선정된 투표자에게 조사를 진행하는 실사 과정에서 발생하는 오차 등으로 구분할 수 있다.

선거예측조사를 통하여 개표 후의 참값과 예측값과의 차이를 비교해 볼 수 있으므로 위에서 언급한 오차들의 방향(+/-)과 크기를 분석할 수 있다. 따라서 본 연구에서는 김영원·곽은선(2010)의 연구에서 사용한 총조사오차(total survey error)개념과 동일한 방식으로 선거구별 예측오차를 구분한다. 즉 전체 투표소 중 일부 투표소를 추출하는 과정에서 발생하는 오차인 ‘투표소추출오차’와, 표본 투표소에서 조사를 수행하는 과정에서 발생하는 계통추출에 따른 표본추출오차와 조사거절이나 거짓응답 등 다양한 원인에 의해 발생하는 오차를 포괄적하는 ‘실사오차’로 구분하기로 한다. 따라서

예측오차는 다음과 같이 두 가지 오차로 구분될 수 있다.

$$\begin{array}{ccccc} \text{예측오차} & = & \text{투표소추출오차} & + & \text{실사오차} \\ (\hat{p}-p) & & (\tilde{p}-p) & & (\hat{p}-\tilde{p}) \end{array}$$

여기서 특정 후보자에 대한 출구조사 예측 득표율은 $\hat{p}$ 이고, 해당 선거구 개표결과 후보자 실제 득표율은 p , 해당 선거구에서 실제 출구조사를 진행한 투표소에 대한 개표결과를 기준으로 산출한 후보자 득표율이 $\tilde{p}$ 이다. 즉, 각 선거구에서 표본 투표소 집합을 s 로 표시하고, 개표결과 i 투표소에서의 후보자의 실제 득표율을 p_i , 실제 투표자 수를 N_i 라고 하면 투표소별 개표결과로 계산한 후보자 득표율 $\tilde{p}$ 는 다음과 같이 계산한 것이다.

$$\tilde{p} = \frac{\sum_{i \in s} N_i p_i}{\sum_{i \in s} N_i}$$

결국 $\tilde{p}$ 는 표본으로 추출된 투표소에서 오차가 없이 완벽한 득표율 추정이 이루어지는 경우 얻어지는 후보자 득표율 추정값이며, 따라서 $(\tilde{p}-p)$ 는 순수하게 투표소 추출에 따른 오차로 볼 수 있다(김영원·곽은선(2010) 참조). 따라서 $(\hat{p}-p)$ 는 출구조사에서 발생한 선거구별 ‘예측오차’를 나타내고, $(\tilde{p}-p)$ 는 해당 선거구에서 투표소 추출 과정에서 발생한 오차로 설명될 수 있는 ‘투표소추출오차’에 해당하고, $(\hat{p}-\tilde{p})$ 는 투표소 내 실사과정에서 발생한 오차를 의미하며, 여기에는 표본 투표소 내 계통추출로 인한 표본추출오차와 함께 추출간격의 미준수, 조사거절이나 거짓응답, 5시 이후 투표자와 부재자 투표 등에 따른 비표본추출오차가 모두 포함되는 것으로 편의상 이를 ‘실사오차’라고 부르기로 한다.

본 연구에서는 새누리당과 통합야당 후보자들에 대한 예측결과를 기준으로 16개 시·도에 대해 각 정당 후보자별 19대 총선 출구조사에서 나타난 예측오차, 투표소추출오차 및 실사오차가 어느 정도 수준인지를 알아보고, 이들 오차에 영향을 미칠 것으로 예상되는 요인들과의 상관분석을 통해 오차의 발생 원인을 살펴보기로 한다. 또한 19대 총선 출구조사를 담당한 세 개 조사기관(MBMR, KRC, TNS)간에 예측오차, 표

본추출오차 및 실사오차가 차이가 있는지를 분산분석을 통해 검증해 보기로 한다.

〈표 2〉는 새누리당 후보자에 대한 16개 시·도별 예측오차, 투표소추출오차 및 실사오차를 정리한 것이다. 서울의 경우 예측오차는 -1.81로 새누리당 후보자의 득표율을 전반적으로 과소 예측하였다. 예측오차를 투표소추출오차와 실사오차로 구분해 보면 투표소추출오차는 -0.2이고 실사오차는 -1.61인 것으로 나타나 결과적으로 예측오차는 대부분 실사오차에서 기인한다는 것을 알 수 있다. 마찬가지로 대구의 경우 투표소추출오차는 0.15인 반면 실사오차는 -1.96으로 나타나 대구지역의 경우에도 실사오차에 의해 새누리당 후보의 득표율을 과소 예측하게 되었음을 알 수 있다.

〈표 2〉 새누리당 16개 시·도별 예측오차

(단위: %)

시·도	예측오차	투표소추출오차	실사오차
서울	-1.81	-0.20	-1.61
부산	-0.56	0.66	-1.22
대구	-1.81	0.15	-1.96
인천	-0.86	0.47	-1.33
광주	0.60	0.19	0.41
대전	-0.68	-0.02	-0.66
울산	-0.93	-0.45	-0.48
경기	-0.68	0.41	-1.10
강원	-0.79	0.36	-1.15
충북	-0.30	0.53	-0.83
충남	0.63	0.38	0.25
전북	1.04	0.31	0.73
전남	-	-	-
경북	0.84	-1.15	1.99
경남	0.62	0.57	0.05
제주	0.11	0.42	-0.31

한편 <표 3>은 통합야당 후보자 득표율에 대한 16개 시·도별 예측오차와 투표소추출오차 및 실사오차를 정리한 것이다. 예측오차를 기준으로 할 때 광주와 충남을 제외한 14개 지역에서 통합야당을 과대 예측하였다.

지역에 따라 과소, 과대 예측의 정도는 다소 차이가 있지만, 오차의 부호를 무시하고 절대 크기만을 비교하였을 때 새누리당은 충남, 경남 및 제주를 제외한 12개 지역에서 투표소추출오차보다 실사오차가 더 컸고, 통합야당은 충남과 경북을 제외한 14개 지역에서 투표소추출오차보다 실사오차가 더 컸다. 이러한 결과는 19대 총선 출구조사에서 발생한 예측오차는 투표소 표본추출과정에 의한 오차보다는 표본 투표소 실사과정에서 조사거절, 추출간격 미준수, 5시 이후 투표자 처리 등의 원인에 의해 발생한 오차가 더 크며, 따라서 출구조사의 예측오차를 줄이기 위해서는 반드시 실사오차를 통제할 수 있어야 한다는 점을 시사한다.

<표 3> 통합야당 16개 시·도별 예측오차

(단위: %)

시·도	예측오차	투표소추출오차	실사오차
서울	2.20	0.37	1.83
부산	1.37	-0.68	2.05
대구	1.86	-0.20	2.06
인천	1.57	-0.18	1.76
광주	-1.36	-0.62	-0.73
대전	1.63	0.38	1.25
울산	1.99	0.96	1.02
경기	1.46	-0.11	1.57
강원	1.77	0.26	1.51
충북	0.83	-0.37	1.20
충남	-1.45	-1.82	0.37
전북	1.93	0.46	1.47
전남	1.95	-0.21	2.16
경북	2.03	1.64	0.39
경남	1.19	-0.41	1.60
제주	1.78	0.39	1.39

IV. 예측오차의 발생 원인 분석

출구조사 후보자별 득표율 예측결과와 실제 개표결과 사이의 오차를 일반적으로 예측오차라고 하며, 예측오차는 앞에서 정의한 바와 같이 투표소추출오차와 실사오차의 합으로 이해할 수 있다. 예측오차에 영향을 주는 요인들로는 선거구별 투표소 수, 총 유권자 수, 총 조사자 수, 부재자투표 비율, 5시 이후 출구조사자 비율, 출구조사 거절률, 투표소 추출률, 선거구내 복합시·군·구 수 등을 고려할 수 있다(이준웅 2012). 이 중 투표소 추출률, 선거구내 복합시군구수는 해당 선거구의 표본추출 난이도와 관련 있는 변수들로 투표소추출오차에 영향을 미칠 것으로 예상되는 요인들이고, 5시 이후 출구조사자 비율과 출구조사 거절률 등은 출구조사 실사를 진행하는 과정에서 발생하는 요인들으로써 실사오차에 영향을 미칠 것으로 예상되는 변인들이다. 여기서 선거구별 총 투표소 수와 부재자 투표율 등은 선관위 자료를 이용하였다(중앙선거관리위원회 2012).

위에서 언급한 요인들과 투표소추출오차 및 실사오차와의 연관성을 알아보기 위해 상관분석을 실시하였다. 여기서는 편향 관점 대신 주요 변인들이 오차의 절대적인 크기에 미치는 영향을 알아보기 위해 모든 오차는 절대 값을 취한 후 상관분석이 이루어졌다는 점을 참고하기 바란다. 먼저 새누리당과 통합야당 후보자에 대한 출구조사 오차와 주요 변인과의 상관분석을 수행한 결과는 <표 4>, <표 5>와 같다. <표 4>와 <표

<표 4> 새누리당 후보자 예측오차와 주요변인과의 상관분석(전국)

	투표소추출오차		실사오차	
	상관계수	p값	상관계수	p값
5시 이후 출구조사자 비율	-0.1602	0.0202	-0.0084	0.9036
총 투표소 수	0.0022	0.9753	0.0523	0.4512
총 조사자 수	-0.2255	0.0010	-0.0969	0.1619
부재자 투표율	0.0847	0.2214	0.0003	0.9972
조사 거절률	-0.0042	0.9514	0.1193	0.0845
투표소추출률	-0.1603	0.0201	-0.1723	0.0124
복합시·군·구 수	0.0140	0.8400	0.0999	0.1493

〈표 5〉 통합야당 후보자 예측오차와 주요변인과의 상관분석(전국)

	투표소추출오차		실사오차	
	상관계수	p값	상관계수	p값
5시 이후 출구조사자 비율	-0.1023	0.1279	0.0152	0.8219
총 투표소 수	0.0107	0.8740	-0.0219	0.7450
총 조사자 수	-0.2322	0.0005	-0.0918	0.1720
부재자 투표율	0.0313	0.6424	-0.0387	0.5652
조사 거절률	-0.0470	0.4857	0.1791	0.0073
투표소추출률	-0.1654	0.0134	-0.0752	0.2634
복합시·군·구 수	0.0088	0.8965	0.0465	0.4901

5)에서 공통적으로 나타난 현상들을 살펴보면, 투표소추출오차의 경우 총 조사자 수와 투표소 추출률이 커지면 투표소추출오차가 감소하는 현상을 볼 수 있는데, 이는 해당 선거구의 경우 상대적으로 다른 선거구에 비해 표본크기가 커지는 효과가 있기 때문에 표본추출오차가 줄어드는 것으로 설명할 수 있다. 실사오차의 경우 조사 거절률이 증가하면 실사오차가 커지는 현상이 새누리당이나 통합야권 후보 모두에서 공통적으로 나타나고 있다. 따라서 분석대상 변수 중 실사오차에 가장 큰 영향을 주는 변수는 조사 거절률이라고 볼 수 있다.

참고로 조성검·김지연(2004) 등이 지적한 것처럼 복합시·군·구 수는 개별 선거구의 표본추출 난이도와 관련 있는 변인에 해당하는데, 예상과는 달리 투표소추출오차에는 큰 영향을 주지 않는 것으로 나타났다. 이런 현상은 현행 출구조사를 위한 표본설계 과정에서 복합선거구 등과 같은 표본추출 난도가 높은 지역에 대해서는 상대적으로 표본 투표소 수를 늘리거나 층화를 통해 표본추출오차를 줄이기 위한 노력을 하고 있고, 이런 표본설계가 긍정적인 영향을 주어 난도가 높은 선거구에 대한 투표소추출오차가 특별히 증가하지 않은 것으로 보인다.

한편 전국 선거구를 대상으로 하면 새누리당과 통합야당 후보 간에 경쟁구도가 이루어지지 못할 수 있고, 영·호남의 경우 지역적인 정치 특성에 따라 출구조사 오차가

영향을 받을 수 있다는 점을 고려해 수도권(서울·경기) 지역의 선거구만을 대상으로 출구조사 오차와 주요 변인들 간의 상관관계를 살펴 본 결과는 <표 6>, <표 7>과 같다. <표 6>과 <표 7>을 보면 특히 실사오차의 경우 조사 거절률과 함께 5시 이후 출구조사자 비율이 증가하면 실사오차가 커진다는 것을 알 수 있다.

<표 6> 새누리당 후보자 예측오차와 주요변인과의 상관분석(수도권)

	투표소추출오차		실사오차	
	상관계수	p값	상관계수	p값
5시 이후 출구조사자 비율	-0.0750	0.4632	0.2532	0.0119
총 투표소 수	0.0223	0.8277	-0.0177	0.8630
총 조사자 수	-0.1751	0.0846	-0.0173	0.8657
부재자 투표율	0.1528	0.1332	-0.0154	0.8802
응답거절률	0.1447	0.1550	0.1807	0.0750
투표소추출률	-0.1224	0.2299	-0.1507	0.1385
복합시·군·구 수	-0.0741	0.4684	-0.0292	0.7754

<표 7> 통합야당 후보자 예측오차와 주요변인과의 상관분석(수도권)

	투표소추출오차		실사오차	
	상관계수	p값	상관계수	p값
5시 이후 출구조사자 비율	-0.0073	0.9427	0.3077	0.0019
총 투표소 수	0.0505	0.6194	-0.0887	0.3828
총 조사자 수	-0.1497	0.1392	-0.0014	0.9890
부재자 투표율	0.1214	0.2312	-0.0104	0.9185
응답거절률	0.0649	0.5235	0.1924	0.0564
투표소추출률	-0.1574	0.1198	-0.0980	0.3346
복합시·군·구 수	0.0365	0.7195	-0.0865	0.3945

출구조사에서 나타난 정당별 예측오차가 16개 시·도 간에 차이가 있다는 것은 <표 2>와 <표 3>에서 살펴보았다. 시·도 간의 차이가 통계적으로 유의하다고 할 수 있는지를 알아보기 위해 분산분석을 실시하였다. 분산분석의 결과는 <표 8>과 같다. 예측오차의 경우 새누리당과 통합야당 모두 16개 시·도 간의 차이가 통계적으로 유의하지 않은 것으로 나타났지만, 투표소추출오차와 실사오차를 구분하여 각각에 대해 분석해보면 새누리당과 통합야당 모두 16개 시·도 간의 차이가 통계적으로 유의한 것으로 나타났다. 이런 현상은 예측오차는 투표소추출오차와 실사오차의 합으로 표현되는데 투표소추출오차와 실사오차의 부호가 다른 경우 실제 예측오차는 두 오차가 서로 상쇄되어 작게 나타날 수 있기 때문이다.

새누리당과 통합야당의 투표소추출오차와 실사오차가 지역에 따라 유의한 차이를 보였다는 것은 지역의 특성이라는 요인이 투표소추출오차 또는 실사오차에 영향을 미쳤다고 볼 수 있으며, 지역에 따라 해당지역의 정치성향에 의해 예측오차를 발생시킬 수 있음을 시사하고 있다. 특히 실사오차 관점에서 보면 지역별 특성에 따라 조사 거절자 처리 등에서 해당 지역에서는 전반적으로 특정 정당 후보에 대해서 일관적으로 일정한 방향의 편향이 발생할 수 있다는 것으로 해석할 수 있다.

다음으로 조사기관별로 차이가 있는지를 알아보기 위해 분산분석을 실시하였다. 19대 총선 방송사 공동출구조사를 담당한 조사기관은 MBMR, KRC, TNS 3개 조사기관으로 246개 선거구를 각 회사별로 82개 선거구씩 나누어 조사하였다. 지역별로는 서울의 경우 48개 표본선거구를 3개 회사가 16개씩 조사하였고, 대구는 표본조사구

〈표 8〉 정당별 예측오차와 시·도(16개 시·도)의 분산분석

		자유도	평균제곱합	F	유의확률
예측오차	새누리당	14	1.92	1.45	0.1343
	통합야당	15	1.66	1.01	0.4451
투표소추출오차	새누리당	14	1.61	2.36	0.0048
	통합야당	15	1.76	2.27	0.0057
실사오차	새누리당	14	2.52	2.19	0.0094
	통합야당	15	2.21	1.80	0.0374

〈표 9〉 조사기관(MBMR/KRC/TNS) 효과에 대한 분산분석

		자유도	평균제곱합	<i>F</i>	유의확률
예측오차	새누리당	2	1.88	1.38	0.2536
	통합야당	2	3.62	2.23	0.1100
투표소추출오차	새누리당	2	0.63	0.85	0.4304
	통합야당	2	1.18	1.41	0.2474
실사오차	새누리당	2	7.57	6.42	0.0020
	통합야당	2	4.85	3.83	0.0233

12개 중 KRC에서 5개, TNS에서 7개 선거구를 각각 조사하였다. 나머지 지역들도 세 개 조사회사가 모두 있는 경우와 두 개 혹은 한 개 회사가 담당한 지역(제주는 KRC만 조사하였음)으로 나눌 수 있다. 조사기관과 예측오차와 투표소추출오차 및 실사오차에 대한 분산분석결과는 〈표 9〉와 같다. 예측오차와 투표소추출오차에 대해서는 통계적으로 유의한 차이를 보이지 않았으나, 실사오차의 경우 새누리당과 통합야당 후보자 모두에 대해 조사기관별 차이가 유의함을 알 수 있다. 특히 투표소추출오차에 있어서 조사기관별 차이가 나타나지 않는 이유는 출구조사 표본설계 과정에서 투표소 추출 방법에 대해서는 KEP 주관으로 일괄적으로 모든 회사가 동일한 방식을 적용하도록 관리하고 있다는 점에서 찾을 수 있을 것이다. 반면 각 조사회사별로 조사원 모집방법이나 실사관리 방식에 있어서 각각 고유의 특징을 갖고 있기 때문에 실사오차에 있어서는 조사회사별로 통계적으로 유의한 차이가 발생한 것으로 보인다.

V. 요약 및 시사점

표본조사 연구자들은 통계조사에서 비표본추출오차가 표본추출오차보다 추정값에 훨씬 더 악영향을 미치는 경우가 많다는 것을 인식하여 왔으며, 비표본추출오차의 원인과 이런 오차를 최소화하는 조사설계와 이런 오차를 줄일 수 있는 추정치 보정방법 등에 대해 많은 관심을 가져왔다(Biemer & Lyberg 2003).

출구조사에서 발생하는 오차 역시 일부 투표소를 추출함으로써 발생하는 투표소추출오차와 투표소 내의 투표자 추출 및 조사거절 등의 다양한 원인에 의해 발생하는 실사오차로 구분해 볼 수 있다. 본 연구에서는 출구조사에서 발생하는 예측오차를 투표소추출오차와 실사오차로 구분하고, 출구조사 오차의 발생 원인과 수준을 알아보기 위해 2012년에 치른 19대 총선 출구조사 자료를 이용하여 예측오차를 분석하였다. 16개 시·도를 구분해 예측오차를 정리해 본 결과 새누리당에 대해서는 과소 예측, 통합야당에 대해서는 과대 예측이 발생하였음을 확인할 수 있었다. 또한 예측오차를 투표소추출오차와 실사오차로 세분하여 분석을 실시하였는데, 대부분의 지역에서 투표소추출오차에 비해 실사오차가 더 크다는 것을 알 수 있었다.

오차를 발생시킬 것으로 예상되는 변인들(투표소 수, 출구조사 조사자 수, 부재자 투표 비율, 5시 이후 출구조사자 비율, 출구조사 거절률, 투표소 추출률, 선거구내 복합시·군·구 수)와 투표소추출오차 및 실사오차와의 상관분석을 실시하였다. 상관분석 결과 투표소 추출률이 증가하면 예상했던 것처럼 투표소추출오차가 감소한다는 것을 확인할 수 있었으며, 5시 이후 출구조사자 비율이나 조사 거절률이 실사오차를 증가시키는 주요 원인이 될 수 있다는 결론을 얻을 수 있었다.

한편 출구조사를 진행한 세 개 조사기관별로 예측오차, 투표소추출오차, 그리고 실사오차의 차이를 분산분석한 결과 특히 실사오차에 있어서 조사기관 간의 차이가 유의한 것으로 나타났다. 따라서 향후 출구조사에서 오차를 줄이기 위해서는 실사오차를 줄이는 방안에 대한 연구를 집중할 필요가 있으며, 그런 관점에서 실사오차를 발생시킬 것으로 예상되는 변수들을 적절히 제어할 수 있다면 출구조사의 정확성을 높이는 데 기여할 수 있을 것으로 판단된다.

참고문헌

- 김영원·김지현. 2007. “2006년 지방선거 출구조사 현황 및 예측오차.” 《조사연구》 8(1): 55-79.
- 김영원·곽은선. 2010. “2008 총선 출구조사의 총조사오차 분석.” 《조사연구》 11(3): 33-55.
- 김영원·엄윤희. 2005. “출구조사를 위한 투표소 확률추출 방법.” 《조사연구》 6(2): 1-32.
- 김영원·최윤정. 2011. “출구조사의 체계적인 예측 편향에 대한 분석: 2010년 지방선거 출구조사

- 를 중심으로.” 《조사연구》12(3): 25-48.
- 조성겸 · 김지연. 2004. “출구조사의 투표소 표집방안 비교.” 《조사연구》 4(1): 3-29.
- 박진우 · 배형민. 2012. “6.2 지방선거 출구조사의 표본크기에 대한 통계적 고찰.” 《조사연구》 13(3): 89-104.
- 이준웅. 2004. “제 17대 총선 예측조사의 문제: 조사 거절자와 응답불성실자 편향을 중심으로.” 《언론정보연구》 41(1): 110-136.
- 이준웅. 2012. “제 19대 총선 출구조사의 교훈: 예측 오차와 편향 영향력 요인” 《2012년도 한국조사연구학회 춘계학술대회 발표 논문집》.
- 중앙선거관리위원회. 2012. 《제19대 국회의원선거총람》
- 허명희. 2008. “국회의원 선거에서의 주요정당 의석 수 예측.” 《조사연구》 9(1): 87-100.
- 홍내리 · 허명희. 2001. “제16대 국회의원 선거의 예측조사에 대한 사후적 검증.” 《조사연구》 2(1): 1-35.
- Biemer Paul P. and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. New Jersey: John Wiley & Sons, Inc.

<접수 2013/10/18 , 수정 2013/11/6 , 게재확정 2013/11/7>

