

연구논문

출구조사의 투표소 표집방안 비교\*

How to Select Polling Places in Exit Poll?

조성겸\*\*, 김지연\*\*\*

Sung Kyum Cho, Ji Yun Kim

출구조사에서 투표소 표집방법으로는 지금까지 판단표집의 일종인 대표구 추출 방법과 층화 대표구의 두 가지 방법이 주로 사용되고 있지만 아직 어느 방법이 더 효과적인가에 대해서 체계적인 연구가 이어지지 않았다. 본 연구는 지금까지 논의 되어 온 네가지 투표소 추출방법 중에서 어느 방법을 사용할 때 오차가 가장 적을 것인지 그리고 어느 정도의 투표소를 추출하는 것이 적절한지에 대해서 분석해 보았다. 17대 총선에서 1위와 2위간의 격차가 8% 이내로 경합을 보인 50개 선거구를 대상으로 투표소 추출방법과 추출규모를 달리했을 때 예측오차가 어떻게 달라지는가를 분석한 결과, 과거 선거결과를 토대로 투표소를 정치성향별로 정렬한 후 체계적으로 추출하는 '종단적 체계적 추출방법'이 지금까지 주로 사용되었던 대표구 혹은 층화후 대표구 방법보다 우수한 것으로 나타났다. 그리고 표집규모는 9개에서 11개 정도의 투표소 추출이 적절한 것으로 나타났다.

주제어: 대표구, 출구조사, 투표소 추출

In Korea, bellwether voting places were selected for exit poll based on the past voting results. Sometimes, voting place stratification were used to improve the exit poll performance. The sampled voting places are intended to mirror the general voters of the entire electoral district. But few studies have been done as to which sampling method works better.

This study compared the four sampling methods—bellwether voting place sampling method, random sampling method, stratified bellwether sampling method and systematic sampling from ordered voting places method. When we applied the four methods to the 2004 general election data, the systematic sampling from ordered voting places method outperformed the other three

\* 이 논문은 2002년도 충남대학교 자체연구비의 지원에 의하여 연구되었음.

\*\* 교신저자(corresponding author): 충남대학교 언론정보학과 교수 조성겸.

E-mail: skcho@cnu.ac.kr

\*\*\* 미디어리서치 사회여론조사본부 본부장.

sampling method. Also, we found that the additional sampling of voting places over nine contribute little to the accuracy of the estimation.

key words: exit poll, bellwether, voting place sampling

## I. 서론

### 1. 문제의 제기

출구조사는 선거예측조사의 대표적인 방법으로 활용되고 있다. 이 방법은 투표자를 대상으로 실시되기 때문에 전화조사와는 달리 투표율이 낮은 상황에서도 정확도가 높을 것으로 기대된다. 전화조사의 경우, 유권자 전체를 모집단으로 삼아 표본을 추출한 후 조사대상자의 응답에 의존하여 투표예상자를 가려내기 때문에 통상적으로 투표율이 낮은 선거에서는 그 정확성이 떨어진다. 이러한 이유에서 우리나라에서는 2000년 총선부터 출구조사 방법이 선거예측에서 본격적으로 활용되고 있고, 미국과 영국 등 외국에서도 출구조사가 주요 선거예측 및 선거분석 조사방법으로 활용되고 있다. 그렇지만 출구조사의 구체적인 절차나 방법에 대해서는 아직 개선할 부분이 많이 있다는 지적이다(류제복 2003; Biemer et al. 2003; Moon 1999). 출구조사의 경우 예측오차는 전화조사보다는 낮게 나타나지만 정당별의 예측에 편향이 나타나는 문제가 발생하고 있다(조성겸 2001).

출구조사 방법에서 특히 논란이 되는 부분은 투표소 표집방법과 표본크기다. 출구조사는 많은 경비가 소요되기 때문에 선거구 내의 투표소 중에서 일부만을 선정하여 조사하는데, 방송사마다 다르기는 하지만 대체로 선거구별로 다섯 개에서 열 개 정도의 투표소를 추출하게 된다. 따라서 투표소를 어떻게 그리고 몇 개를 선정해야 하는가의 문제가 중요해진다. 추출 투표소 수를 증가시키면 예산도 증가될 뿐만 아니라 필요한 조사원을 모집해서 배치하는 것도 어려운 문제다<sup>1)</sup>.

현재까지 투표소 추출은 일종의 판단표집(judgement sampling)방법을 이용해서 이루어졌다. 이러한 방법 중의 하나가 '대표구(bellwether polling place) 추출방법'인데 이것은 선거구의 투표성향을 가장 잘 대표할 수 있다고 판단되는 투표소를 선정하여 조사하는 방법이다. 투표소 추출에서 확률표집(probability sampling) 방법보다는 판단표집방법을 더 많이 사용하게 된 것은 확률표집방법이 조사의 효율성 측면에서 문제가 크다고 보았기 때문이다. 확률표집방법을 사용할 경우 표집오차를 적절한 수준으로 낮추기 위해서는 상당한 수의 투표소를 추출해야 하지만 이렇게 되면 면접원 수와 비용이 증가하게 된다. 반면에 판단표집방법은 가장 대표적인 투표소를 추출하는 것이기 때문에 그 수가 많아야 할 이유는 없다고 보았다. 즉 실제 선거조사에서 5개에서 9개 정도의 투표소를 선거구별로 추출하게 되는데 이 정도의 표본크기라면 확률표집보다는 판단표집방법이 더 나을 것이라고 보았기 때문이다.

한국 선거의 경합도는 높은 편이다. 1997년 15대 대선의 경우 1,2위 후보자의 득표율 차이가 1.6%였고, 2002년 16대 대선은 2.3%였다. 그리고 2000년 16대 총선의 경우 20표 이내에서 승부가 갈린 서울 동대문乙, 경북 울진/봉화, 경기 광주 등을 포함하여 1% 이내가 14개였으며, 5% 이내의 경합지역이 53개에 달하고 있다. 또한 2004년 17대 총선 역시 9표차를 보인 충남 당진을 비롯하여 1% 이내가 8개였으며, 5% 이내의 경합지역이 43개에 이르렀다. 따라서 출구조사를 이용한 예측이 실패하지 않기 위해서는 투표소 추출과정에서 발생하는 오차를 낮추어야 한다.

이에 따라 그동안 방송사의 출구조사 설계과정에서는 투표소 추출 방법에 대한 많은 논의가 이루어졌다. 초기에는 대표구 표집방법 한

1) 예컨대 115개 선거구에서 평균 8개의 투표소를 추출하고 각 투표소별로 4명의 면접원을 배치할 경우 3,680명의 조사원이 필요하게 된다.

가지만을 이용했지만 층화 대표구 등 변형된 방법들이 활용되었고, 또 종단적 계통표집방법 등도 논의되었다. 그렇지만 각 방법이 갖는 장점과 단점이 분명하지 않은 상태에서 선뜻 실제 출구조사에 적용하기 어렵기 때문에 방법적인 변화가 많이 이루어지지 못했다.

본 연구는 이러한 맥락에서 출구조사에서 현재 사용되고 있는 판단표집 방안과 아직 사용되지는 않았지만 대안적 방법으로 검토되었던 방안을 경험적 자료를 토대로 비교·검토함으로써 투표소 표집방법 선택에 도움이 되고자 한다. 이렇게 함으로써 본 연구는 출구조사방법의 개선은 물론, 유권자의 투표행태에 관한 연구에도 도움이 될 수 있을 것으로 기대한다. 과거 투표행태를 토대로 선정한 투표소가 그 특성을 어느 정도 유지하는가는 곧 투표행태에 영향을 주는 결정요인의 변화 정도를 나타내 준다고 볼 수 있기 때문이다.

## 2. 연구문제

한국에서 출구조사를 위한 투표소를 추출할 때, 과거 선거결과를 토대로 선거구를 대표할 수 있는 투표소를 추출하는 대표구 추출방법, 그리고 과거 선거결과에 따라 층을 구분한 후 층내 대표 투표소를 추출하는 층화 대표구 추출방법의 두 가지를 가장 많이 사용해 왔다. 그렇지만 선거구에 따라서는 투표소가 과거 선거 때와 비교했을 때 많이 달라지는 경우도 있다. 이럴 경우에는 과거 선거결과를 활용할 수 없기 때문에 단순 계통 표집방법 등 확률표집방법을 사용해 왔다. 이 외에도 아직 실제 적용되지는 않았지만 과거 선거결과에 따라 투표소를 배열한 다음 계통적으로 추출하는 방안, 즉 종단적 계통 표집방법이 있다. 이 방법 역시 투표소 변화가 많을 경우에 적용하기는 문제가 있다고 판단된다.

지금까지 실제 출구조사 과정에서 사용되어 왔던 투표소 추출방법 중 어느 방법이 더 효율성이 높은지 그리고 대안적 방법들은 과연 어

떠한지에 대해서는 아직 체계적으로 검토되지 않았다. 구체적으로 확률표집방법의 경우, 투표소 수가 많아야 한다는 이유로 극히 제한된 조건에서만 사용되었지만, 실제로 그러한 이유가 타당한가에 대한 경험적인 검토는 없었다. 따라서 이 연구는 투표소 추출방법을 실제 과거 선거자료에 입각해서 체계적으로 검토해 봄으로써 각 방법의 효율성을 비교해 보고자 한다. 그럼으로써 출구조사의 정확도를 높이는 데 기여함과 동시에 아울러 투표행태에 관련된 연구에 도움이 되고자 한다. 구체적으로 본 연구는 다음과 같은 연구문제를 설정하고 이에 대한 답을 구해 보았다.

- 연구문제 1: 출구조사 투표소 추출의 방법별로 어떤 방법의 정확도가 가장 높은가?
- 연구문제 2: 출구조사의 적절한 투표소 수는 얼마인가?

## II. 투표소 추출방법에 대한 검토

### 1. 단순 계통 추출방법(systematic sampling from random population)

투표소 추출에 활용되는 확률표집방법 중 가장 대표적으로 사용하고 있는 것이 단순 계통 추출방법이다. 표집 틀(sampling frame)로 사용하는 투표소별 유권자 수 정보가 선관위를 통해 공고되기 때문에 계통 추출방법을 적용하기에 어려움은 없다. 구체적 추출절차에 따라 여러 가지 방법이 있을 수 있지만, 투표소별 크기를 고려해서 추출하는 방법과 고려하지 않는 방법 두 가지로 크게 구분할 수 있다. 한국의 출구조사에서는 추출된 투표소에서 응답자를 추출할 때 투표소 크기에 비례해서 추출하게 된다<sup>2)</sup>. 따라서 투표소 추출단계에서는 투표소 크기

2) 투표소에서 조사대상자를 추출할 경우 5명에서 7명 정도의 간격으로 계통 추출을 하게 된다. 따라서 투표소 크기에 따라 추출된 표본의 크기가 달라지게 된다.

를 고려할 필요가 없다. 투표소는 그 크기에 관계없이 동일 확률로 추출하고, 추출된 투표소에서 표집하는 응답자 수를 투표소 크기에 비례하도록 하는 방법이 대표적으로 사용된다.

출구조사의 표집오차는 추출되는 투표소 수 그리고 급내 상관계수 (intraclass correlation)에 따라 달라지지만, 실제 출구조사에서 어느 정도의 표집오차가 발생하는지에 대해서는 아직 명확한 분석이 이루어지지 않았다. 그러나 김재광(2004)의 분석은 대략적인 오차의 크기를 짐작하게 하는데, 그는 16대 총선거 자료를 토대로 16개 선거구의 급내 상관계수를 산출한 다음 그것을 토대로 선거구별 예측치의 오차한계가 대략 3.51%에서 9.83%정도일 것으로 추정하였다. 이 때 각 선거구별로 10개의 투표소를 추출하고 각 투표소별로는 200명의 응답자를 표집하는 것으로 가정하였는데 이것은 실제 17대 총선에서 KBS와 SBS의 출구조사 규모와 비슷한 수준이다. 따라서 이 연구결과가 17대 총선 출구조사에도 적용될 수 있다고 본다.

이렇게 볼 경우, 1위와 2위 후보자간 경합도가 높은 총선의 경우 출구조사가 순위를 정확하게 예측하기 위해서는 위에서 가정한 선거구별 10개보다 훨씬 많은 투표소를 추출해야 한다는 결론이 나오게 된다. 비용대비 효율성이 떨어지기 때문에 단순 계통 추출방법은 투표소 구획의 변화가 심하여 과거 선거자료를 활용할 수 없는 경우에 한하여 적용하고 있다. 예컨대 17대 총선 KBS-SBS 공동 예측조사의 경우 전체 출구조사 대상지역 115개 선거구 중에서 30개 선거구에서 단순 계통 추출방법을 사용하였다.

## 2. 대표구 추출방법(bellwether polling place sampling)

확률표집방법을 사용할 경우, 통계적으로 안정적인 결과를 얻기 위해서는 많은 수의 투표소를 추출해야 하는 것으로 믿어 왔기 때문에 한국의 출구조사에서는 이보다는 다소 불확실한 방법이지만 대표적인

투표소를 소수 선정하는 대표구 추출방법을 채택하여 왔다. 이 방법은 선거구 전체의 투표결과를 대표할 수 있을 것으로 판단되는 투표소를 추출하여 조사하는 방법이다. 대표 투표소를 잘 고를 경우, 그 투표소의 투표결과만으로 전체 선거구의 투표결과를 알 수 있다고 가정하는 것이다. 대표 투표소를 찾는 방법은 여러 가지가 있을 수 있다. 예컨대 소득이나 연령, 종교 등 주요 유권자 특징이 전체 선거구를 대표할 수 있는 선거구를 찾을 수도 있을 것이다. 그러나 출구조사에서는 과거 선거에서의 투표소별 투표결과를 보고, 선거구 전체의 투표결과와 가장 비슷한 결과를 보인 투표소를 대표 투표소로 선정하게 된다. 유사성 정도는 개별 투표소의 후보별 득표율과 선거구 전체의 후보별 득표율 차이를 가지고 판단하는데, 다음 식에서 보듯이 각 후보별 득표율 차이를 제곱하여 그 총합이 적을수록 유사성이 높다고 본다.

$$A \text{ 투표소의 오차합} = \sum_{k=1}^n (X_{.k} - x_{Ak})^2$$

$x_{.k}$  : k위 득표자의 선거구 전체 득표율

$x_{Ak}$  : k위 득표자의 A투표소 득표율

$k = 1, 2, \dots, n$ 위

이와 같이 과거 선거에서 선거구 전체의 투표결과를 대표할 수 있는 투표소를 찾게 되면 이 투표소가 이번 선거에서도 그러한 대표성을 유지할 것으로 보는 것이다. 이 때, 선거구 전체를 가장 잘 대표할 수 있는 투표소가 존재한다면 그 투표소 하나만을 추출하면 될 것이다. 그렇지만 실제로 그러한 대표성에 확신이 없기 때문에 전체 선거구 결과와 가장 유사한 결과를 보인 5~9개 정도의 투표소를 추출하게 된다. 이러한 이유로 투표소 수가 많아진다고 해서 더 정확해진다고 보기는 어렵다.

문제는 투표소가 가진 대표성이 과연 계속 유지될 수 있는지는 것이다. 미국의 1916년 이후 선거자료를 토대로 하여 이러한 대표구 (bellwethers)가 존재하는지 검토한 연구(Tufte & Sun 1975)에 의하면 이러한 대표구는 존재하지 않는다고 한다. 결국 투표행태는 변화한다는 것이다. 그렇지만 앞서 지적하였듯이 적은 수의 투표소를 가지고 예측해야 한다면 확률표집방법보다는 이 방법이 더 정확할 것으로 예상되었다. 특히 유권자 유입과 유출에 따른 인구변동이 적고, 선거구 조정에 따른 급격한 정치환경 변화가 없는 경우에는 특히 이러한 대표구 방식이 효율적인 것으로 보고 있다. 이러한 믿음은 총선 이후 시간이 그리 경과되지 않은 보궐선거 등에서, 즉 투표소 변동이 적은 선거에서 이러한 방법을 이용하여 투표소를 추출한 다음 출구조사를 실시하였을 때 그 결과가 비교적 정확했다는 경험에 의해 뒷받침되고 있다.

### 3. 층화 대표구 추출방법

(bellwether polling place sampling after stratification)

대표구 추출방법이 선거구 전체의 결과를 대표할 수 있는 투표소를 추출하는 것이라면 층화 대표구 방법은 선거구를 몇 개의 층으로 구분한 다음 각 층별로 대표 투표소를 찾는 방법이다. 보통 선거구를 2개에서 4개 정도의 층으로 구분한 다음 각 층을 대표할 수 있는 투표소를 층별로 두 개 이상 추출한다. 이 방법은 선거구에 어떤 변화가 있더라도 층별로 구분하여 표집하기 때문에 이러한 변화가 표집에 반영될 수 있다고 본다. 그렇지만 층 구분을 어떻게 해야 하는지와 관련한 방법의 선택이나 층 수의 결정이 연구자에 따라 달라지기 때문에 실제 적용과정에서 어려움이 많다. 이 방법은 KBS의 2002년 지방선거 예측에서 활용되어 비교적 성공적이라는 평가를 받았다. 당시 출구조사에서는 투표소의 변화를 대표구 방식이 적절하게 반영하지 못한다고 보았다. 따라서 층별로 구분해서 대표구를 추출하는 것이 보다 적합할



것으로 보고, 각 선거구를 2개에서 4개 정도의 층으로 구분한 다음, 각 층별로 적절한 수의 선거구를 표집하여 예측치를 구하였는데, 이때의 출구조사는 당선자 예측이 빗나가지 않았고, 실제 예측오차도 그리 크지 않았다. 그리고 사후 분석결과 투표소 추출과정에서 발생한 오차의 평균은 1%포인트 정도인 것으로 밝혀졌다. 따라서 그 이후 2002년 보궐선거와 대통령 선거에도 이 방법이 사용되었다. 2004년 총선에서는 복합선거구 등 일부 선거구에서 사용되었다.

#### 4. 종단적 계통 추출방법

(longitudinal systematic sampling from ordered polling places)

선거구 전체의 대표 투표소를 찾는 방법은 투표소에 발생하는 변화가 반영되지 않기 때문에 한정적이지만 이를 극복하려는 층화 대표구는 층을 구분하는 방식에서 주관성이 많이 개입된다는 문제점이 있다. 이러한 문제점을 극복할 수 있는 방법으로 종단적 계통 추출방법을 들 수 있다. 이 방법은 특정 정당이나 후보자의 득표율을 기준으로 투표소를 배열한 다음 일정한 간격으로 즉 계통적으로 투표소를 추출하게 된다. 대표구 추출방법과 다른 점은 투표성향이 각기 다른 층에서도 투표소를 추출하기 때문에 선거구의 변화를 반영해 준다는 점이다. 즉, 대표구 방식은 선거구 전체의 평균과 가장 유사한 투표성향을 보이는 투표소들만을 뽑는 반면, 종단적 계통 추출방법은 여당지지가 강한 투표소, 여야의 지지율이 비슷한 투표소, 야당지지가 강한 투표소 등을 모두 뽑는다는 특징이 있다. 그러면서도 층을 구분하지 않기 때문에 층화 대표구 방법의 단점인 층 구분의 임의성이 배제된다는 장점도 있다. 그렇지만 층이 명확하게 구분되어 있을 때는 경우에 따라 층별 특성이 적절하게 반영되지 않을 위험성도 있다. 예컨대 비슷한 크기의 두 개 층으로 구분되는 선거구에서 세 개의 투표소를 추출할 경우 층의 크기는 유사하지만 추출된 투표소 수는 달라진다. 만약 각 층별

로 성향차이가 크고 층내의 성향차이가 적다면 층 구분을 무시하고 표집하는 종단적 계통 표집방법이 부정확하게 될 수도 있을 것이다.

이 방법은 우리나라에서는 아직 적용하지 않았지만 영국의 출구조사에서 활용되었는데(Moon 1999), 이들은 이 방법을 그대로 사용할 경우 갖게 되는 편향을 제거하기 위해 대칭추출방식(mirror-image)을 이용하여 정치적 성향의 균형을 꾀하도록 변형하여 사용하였다. 우리나라에서 이 방법이 활용되지 않았던 것은 그동안 대표구 방법에 대한 확신이 강했기 때문에 굳이 새로운 방법을 도입할 필요성을 느끼지 못했기 때문일 것으로 판단된다.

### III. 연구방법

한국에서는 1999년 3.30재보선에서 최초로 출구조사가 도입된 이후, 1999년 6.3재보선, 2000년 16대 총선, 2002년 3회 지방선거, 2002년 8.8재보선, 2002년 16대 대선, 그리고 최근의 2004년 17대 총선에 이르기까지 총 7번의 출구조사가 이루어졌다. 이 중 16대와 17대 총선에서 출구조사의 정확도가 떨어지는 것으로 나타났고, 다른 선거에서는 비교적 정확한 예측이 이루어졌다. 투표소 표집이 특히 문제가 되는 경우는 총선이기 때문에 본 연구는 17대 총선 자료를 이용하여 표집방법을 비교하였다. 총선 출구조사는 선거구 수가 많기 때문에 선거구별 투표소 수를 많이 추출할 수 없고 또 예측이 빗나가는 경우도 많았기 때문이다.

#### 1. 투표소 추출에 사용된 자료

대표구 표집방법, 층화 대표구 표집방법, 종단적 계통 표집방법은 모두 과거 투표행태를 기준으로 투표소를 표집하는 방법이다. 17대 총선 예측조사에서 사용할 수 있는 과거 투표행태 자료로는 2000년에 실

시된 16대 총선 결과와 2002년에 실시된 16대 대선 결과 등이 있다.

총선의 경우 선거구의 투표소별 투표행태가 대통령 선거나 지방선거와는 다를 것으로 보인다. 예컨대 대통령 선거나 광역단체장 선거의 경우, 후보구도와 이에 따른 선거테마가 전국을 아우르기 때문에 개별 선거구 단위의 특성들은 중요하게 인식되지 않는다. 반면, 총선의 경우에는 이러한 요인 이외에도 개별 후보자의 출신지역, 출신학교 등이 투표행태에 중요한 영향을 미치게 되는 등 다른 선거와 차이가 있다. 따라서 총선자료를 이용한 대표구 추출을 실시하는 것이 이론적으로는 가장 적합하다. 그러나 17대 총선의 경우 16대에 비해 지역구 국회의원 수가 227개에서 243개로 늘어나면서 선거구 구역조정이 일어난 곳이 <표 1>에서 보는 바와 같이 75개에 달한다. 따라서 16대나 그 이전 선거행태를 토대로 17대 총선에서의 투표소별 특성을 파악하는 것은 상당한 어려움이 있다.

<표 1> 16대 총선 대비 17대 총선 선거구 변동 현황

유 형	선거구
1개 선거구가 2개로 분리	서울 성동 갑/을 등 44개
3개 선거구가 2개로 통합	전북 김제/완주, 진안/무주/장수/임실 등 10개
2개 선거구가 3개로 분리	서울 노원 갑/을/병 등 9개
2개 선거구가 1개로 통합	대구 남구/중구 등 4개
복합선거구 소속 시/군의 변화	강원 화천/철원/양구/인제 등 4개
신설 선거구	경기 수원 영통 등 1개
신설 선거구에 따른 구역조정	경기 수원 팔달 등 3개
계	75개

이 연구는 2000년 16대 총선이 아닌 2002년 16대 대선 투표자료를 투표소 표집을 위한 자료로 활용하였다. 실제로 지난 17대 총선 출구 조사에서도 이러한 이유로 투표소 표집과정에서 16대 대선 자료를 이용하였다.

## 2. 분석대상 지역

실제 선거예측조사에서 출구조사는 1위와 2위 후보자간의 예상 득표율 차이가 8% 이내일 경우에 실시되는 경우가 많았기 때문에 본 연구에서도 8%를 기준으로 하여 실험지역을 정하였다. <표 2>에서 살펴볼 수 있듯이 17대 총선에서는 243개 지역구 중에서 80개 지역이 8% 이내의 경합도를 보였다.

<표 2> 17대 총선 1~2위 격차가 8% 이하인 선거구의 16대 대선 대비 투표소 수 증감 현황

0	±1	±2	±3 이상	계
20개 (25.0%)	11개 (13.8%)	7개 (8.8%)	42개 (52.5%)	80개 (100.0%)

실제 분석대상이 된 지역은 38개였는데 그것은 이들 80개 지역 중에서 16대 대선에 비해 투표소 변동이 3개 이상 발생한 지역은 제외했기 때문이다. 투표소 변동이 이처럼 많을 경우 과거 투표행태를 기준으로 투표소를 추출하지 못하고 계통추출법과 같은 확률표집방법을 사용할 수밖에 없다. 따라서 방법별 비교라는 연구목적상 17대 총선에서 경합도가 높으면서도 투표소에 변동이 적었던 38개 선거구를 대상으로 투표소 표집방법을 비교하였다.

〈표 3〉 분석에 투입된 선거구 현황

No	시도	선거구	투표소 수 증감	No	시도	선거구	투표소 수 증감	No	시도	선거구	투표소 수 증감
1	서울	종로	0	14	인천	남동乙	0	27	서울	영등포甲	1
2	서울	용산	0	15	인천	부평甲	0	28	서울	강동甲	-1
3	서울	광진乙	0	16	울산	울주군	0	29	부산	영도	1
4	서울	은평乙	0	17	경기	남양주甲	0	30	부산	북강서甲	1
5	서울	서대문乙	0	18	경기	파주	0	31	부산	사하乙	1
6	서울	양천乙	0	19	전북	고창/부안	0	32	경기	광명乙	-1
7	서울	강서乙	0	20	경남	양산	0	33	경기	고양시일산甲	2
8	서울	영등포乙	0	21	서울	동대문甲	1	34	경기	남양주乙	2
9	서울	동작甲	0	22	서울	노원乙	-2	35	강원	춘천	-2
10	서울	송파乙	0	23	서울	서대문甲	1	36	강원	철원/화천/양구/인제	-2
11	서울	송파丙	0	24	서울	마포甲	-1	37	전남	담양/곡성/장성	-2
12	서울	강동乙	0	25	서울	마포乙	-2	38	전남	고흥/보성	-1
13	인천	남구甲	0	26	서울	금천	1				

### 3. 구체적인 표집절차

분석을 위해 네가지 표집방안별로 표본크기(투표소 수)를 5, 7, 9, 11, 13개로 달리하여 총 20개 경우의 투표소 추출을 실시하였다.

추출 투표소 수를 5~13개로 삼은 것은 실제 방송사 출구조사도 그 정도 범위 내에서 진행되기 때문이다. 총선을 기준으로 선거구별 투표소는 대략 50개에서 60개 정도<sup>3)</sup>이며, 이 중에서 7~10개 전후의 투표소가 추출되어 최종예측에 사용된다. 지난 몇 차례의 출구조사를 통해 비용 대비 효율성 측면에서 이 정도의 표본크기가 적절하다고 판단되었기 때문이다. 실제 분석에서 사용한 각 표집방법별 구체적 절차는

3) 17대 총선의 전체 243개 선거구를 분석한 결과 1개 선거구당 평균 54.2개 투표소로 구성되어 있다.

다음과 같다.

### 1) 단순 계통 추출방법

단순 계통 추출방법에 의한 투표소 추출을 다음과 같은 절차에 의해 시행하였다. 먼저, 17대 총선 자료를 기준으로 선거구 내에서 투표소를 무작위로 배열한 후, 난수를 발생시켜 최초 추출점(starting point)에 해당되는 투표소를 추출하였다. 다음으로 배열된 투표소 전체 수를 추출 투표구 수(k)로 나누어 간격(interval)을 구한 다음, 그 간격에 따라 계통적으로 건너뛰며 해당하는 투표구를 선정하였다.

### 2) 대표구 추출

대표구 추출방법은 선거구 전체의 투표결과를 대표할 수 있는 투표소를 뽑는 방법인데, 구체적인 방법은 여러 가지가 있다. 우선 1단계에서 투표소를 바로 뽑는 일단추출법과, 상위 행정단위인 읍/면/동, 혹은 시/군/구를 먼저 뽑고 그 다음에 투표소를 뽑는 다단추출법으로 나뉘지는데, 본 분석에서는 일단추출법을 사용하였다.

실제 추출하는 과정은 16대 대선 자료를 이용하여 선거구 전체의 후보별 지지율에서 개별 투표소의 후보별 지지율을 뺀 값을 제공하여 합한 다음, 그 크기에 따라 정렬한 후 그 차이가 적은 순서대로 k개의 투표소를 뽑았다. 이러한 이유 때문에 대표구 표집방법이 오차 최소제곱법(least squares method)이라고도 불린다.

### 3) 층화 대표구 추출

선거구를 몇 개의 층으로 구분한 다음 각 층별로 대표구를 추출하는 방법이다. 이 방법의 성과는 층을 어떻게 나누느냐에 달려 있는데 그 방법에 따라 오차가 커지기도, 줄어들기도 한다. 층내 동질적이며 층간 이질적이라는 기본 원칙이 잘 반영되게 층을 나누는 것이 중요하다. 층을 나누는 방법은 투표소별 각 후보의 과거 선거 득표율을 토대로 군집분석 등 통계 모듈을 적용하는 방법과 각 투표소별로 지지성향

의 점수 값을 부여하고 이 분포를 보고 연구자의 판단에 의존하여 층을 나누는 방법이 있다. 본 연구에서는 개별 투표소 단위의 16대 대선 후보별 지지율 자료를 가지고 군집분석을 실시하여 층을 2개~5개로 구분하였으며, 층을 나눈 후는 앞의 대표구 표집방법과 동일한 단계를 거쳤다.

#### 4) 종단적 계통추출

선거구를 층으로 구분하지 않고, 선거구 내의 투표소를 투표성향별로 나열한 다음 체계적으로 추출하자는 것이다. 먼저, 16대 대선 자료를 이용하여 선거구 내에서 노무현 후보 지지율 순서대로 투표소를 정렬하였다. 그 다음 단계부터는 앞의 단순 계통 추출방법을 그대로 사용하였다. 단순 계통 추출의 경우 선거관리위원회의 투표소 배열을 무작위적인 것으로 전제하고 있다면, 종단적 계통 표집은 과거 선거의 여야 지지성향을 기준으로 투표소를 재정렬한 후 추출한다는 차이가 있다.

### 4. 투표소 추출오차(예측오차)

투표소 추출방법별 비교는 각 방법을 사용했을 때의 예측오차<sup>4)</sup>를 산출하여 이를 토대로 실시하였다. 예측오차를 구하는 방법으로는 당선자만을 기준으로 구하는 방법, 1위와 2위 후보자 두 명의 오차를 평균하거나 더하는 방법, 모든 후보자의 오차를 합하는 방법이 있다. 그러나 선거예측에서는 각 후보자의 득표율 예측의 정확도보다는 상대적 정확도가 더 중요하기 때문에 1위와 2위 후보의 득표율 격차를 기준으로 다음과 같이 예측오차를 구하였다.

$$\text{예측오차} = (1\text{위 득표자의 선거구 전체 득표율} - 2\text{위 득표자의 선거구 전체 득표율}) - \\ (1\text{위 득표자의 추출된 투표소 득표율} - 2\text{위 득표자의 추출된 투표소 득표율})$$

4) 선거구 전체결과와 추출투표소 결과의 차이

## IV. 연구결과

출구조사 표본의 추출방법을 단순 계통, 대표구, 층화 대표구, 종단적 계통 방법으로 각기 달리하였을 때 예측오차가 어떻게 달라지는가를 분석하기 위해 17대 총선거자료를 이용해 38개 선거구를 대상으로 표본을 추출한 다음 예측오차를 산출하였다. 또한, 추출 투표소 수의 차이에 따른 예측오차의 변화도 함께 검증하기 위하여 각 선거구별로 표본크기(추출 투표소 수)를 5,7,9,11,13으로 각기 달리하여 분석하였다. 그 결과 네가지 표집방법별로 5개의 표본크기별 예측오차가 계산되어 총 20개의 각기 다른 예측결과가 산출되었다. 그 다음 이러한 예측결과를 실제 투표결과와 비교하여 예측오차를 구하였다. 그 다음 이 예측오차가 추출방법 및 추출 투표소 수에 따라 어떻게 달라지는가를 분석해 보았다.

### 1. 투표소 수에 따른 예측오차 차이

투표소 수를 증가시키면 예측오차는 감소하지만 예산과 실행의 어려움이 있기 때문에 적정 투표소 수를 추출하는 것이 중요하다. 그동안 총선 출구조사에서는 선거구별로 5개에서 13개 정도의 투표소를 추출해 조사했지만 과연 어느 정도의 투표소를 추출하는 것이 적절한 것인가에 대한 체계적인 분석이 이루어지지는 않았다. 본 연구는 투표소 수를 달리했을 때 예측오차가 어떻게 달라지는가를 주요 연구문제로 설정하고 이를 분석하였다.

분석결과 <표 4>에서 보듯이 투표소 수를 증가하면 예측오차가 감소하는 경향이 있었다. 단순 계통 표집의 경우, 대체적으로 투표소 수가 증가할수록 예측오차가 작아지는 경향을 보였다. 대표구 표집의 경우는 5개 추출에 비해 7개를 추출하였을 때 예측오차가 크게 감소하였으며, 그보다 투표소 수가 증가하더라도 오차가 거의 감소하지는 않았



다. 층화 대표구 방식으로 표집하였을 경우는 9개와 11개가 상대적으로 우수한 편이었으며, 종단적 계통표집의 경우는 투표소 수와 상관없이 대체적으로 높은 예측력을 보였다.

〈표 4〉 표집방법 및 투표소 수별 예측오차

표집방법	투표소 수	평균	표준오차
단순 계통 표집	5개	2.683	0.372
	7개	2.784	0.404
	9개	1.926	0.282
	11개	2.457	0.285
	13개	1.861	0.245
대표구 표집	5개	2.672	0.655
	7개	1.892	0.463
	9개	1.741	0.379
	11개	1.663	0.331
	13개	1.760	0.381
층화 대표구 표집	5개	2.793	0.422
	7개	2.281	0.333
	9개	1.912	0.337
	11개	1.950	0.345
	13개	2.204	0.389
종단적 계통 표집	5개	1.215	0.128
	7개	1.925	0.267
	9개	1.428	0.271
	11개	1.169	0.211
	13개	1.414	0.291

이러한 예측오차가 표집방법과 투표소 수에 따라 과연 달라지는가를 알아보기 위해 SPSS의 GLM 모듈을 이용해 분석하였는데 그 결과 투표소 수의 영향은 통계적으로 의미있는 것으로 나타났다.

〈표 5〉 변량 분석 결과

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
표집방법(a)	Huynh-Feldt	94	2	55	3.9	.030
error	Huynh-Feldt	888	64	14		
투표소 수(b)	Huynh-Feldt	45	3	16	4.9	.004
error	Huynh-Feldt	342	104	3		
상호작용(a*b)	Huynh-Feldt	41	8	5	1.5	.176
error	Huynh-Feldt	1044	278	4		

\* SPSS GLM 모듈의 repeated measures design 사용

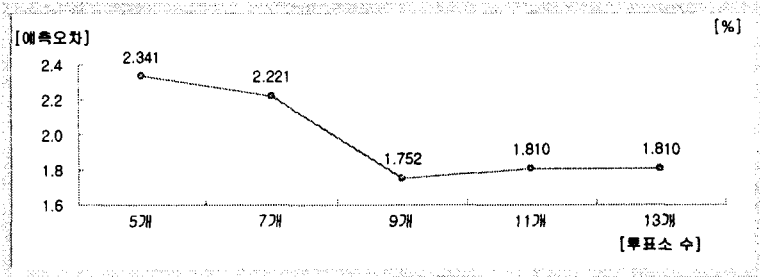
그리고 추출 투표소 수를 증가시킬 경우, 예측오차가 어떻게 변화하는가를 구체적으로 측정하기 위해 세부적으로 분석해 본 결과 〈그림 1〉과 〈표 6〉에서 살펴볼 수 있듯이 투표소 수를 9개로 할 때까지 오차가 감소하였지만 9개를 초과하는 투표소 수 추출은 예측오차의 감소에 기여도가 별로 없는 것으로 나타났다. 이 결과는 그동안 7~10개 정도의 투표소를 추출하는 것이 적절하다는 그동안의 가정에 큰 문제가 없음을 보여 준다.

〈표 6〉 투표소 수 Contrast 분석 결과

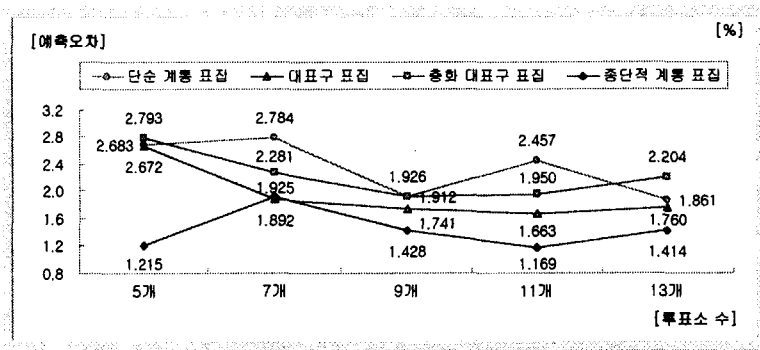
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
투표소 수	Level 2 vs. Level 1	1	1	1	0.4	.511
	Level 3 vs. Previous	11	1	11	14.8	.000
	Level 4 vs. Previous	3	1	3	2.9	.966
	Level 5 vs. Previous	2	1	2	4.1	.051

\* Level 1: 5개 추출, Level 2: 7개 추출, Level 3: 9개 추출, Level 4: 11개 추출, Level 5: 13개 추출

가. 전체



나. 방법별



<그림 1> 투표소 수 증가에 따른 예측오차 변화

2. 투표소 추출방법에 따른 예측오차 차이

다음으로 표집방법별 예측오차 차이분석을 실시하였다. 본 단계에서는 투표소 수 9개 이상만을 분석대상으로 삼았다. 앞에서 살펴본 바와 같이 투표소 수 5개와 7개는 표집방법과 상관없이 대체적으로 예측오차가 컸는데, 이러한 이유로 표집방법별 변별력을 떨어뜨릴 것으로 판단했기 때문이다.

분석 결과, 표집방법의 유의확률이 .061로 나타나 95%신뢰수준에서는 표집방법별 유의미한 차이가 인정되지 않았다. <표 8>의 각 방법

별 다중 비교를 살펴보면, 종단적 계통 표집의 예측오차가 단순 계통 표집보다 유의미하게 작았지만, 다른 조합의 경우에는 통계적인 차이가 발견되지 않았다.

<표 7> 표집방법 변량 분석 결과

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
표집방법 Huynh-Feldt	40	2	21	3.0	.061
error Huynh-Feldt	497	71	7		

\* 투표소 수 5개와 7개는 분석에서 제외

<표 8> 표집방법에 따른 다중 비교

(I) 표집방법		(J) 표집방법	평균차 (I-J)	표준오차	유의확률
Tukey HSD	단순계통	대표구	0.360	0.338	0.876
		총화대표구	0.059	0.244	1.000
		종단적계통	0.744	0.199	0.004
	대표구	단순계통	-0.360	0.338	0.876
		총화대표구	-0.301	0.215	0.671
		종단적계통	0.384	0.374	0.894
	총화대표구	단순계통	-0.059	0.244	1.000
		대표구	0.301	0.215	0.671
		종단적계통	0.685	0.268	0.086
	종단적계통	단순계통	-0.744	0.199	0.004
		대표구	-0.384	0.374	0.894
		총화대표구	-0.685	0.268	0.086

\* 투표소 수 5개와 7개는 분석에서 제외

당선자 예측이 가장 중요하게 고려되는 실제 선거예측 과정에서는 예측오차의 평균보다는 오차의 분산이 표집방법 선택에 있어서 더 중요하게 고려되기도 한다. 즉, 예측오차평균이 적은 표집방법이라 하더라도

라도 분산이 클 경우, 보다 정확히 말하자면 극단치의 수가 많을 경우, 당선자 예측실패 선거구 수가 늘어나기 때문이다. 반면, 오차평균이 상대적으로 떨어지는 방법론이라 하더라도 극단치의 수가 적다면, 절대다수의 선거구에서 당선자 예측에 성공할 수 있기 때문이다.

〈표 9〉를 살펴보면, 일단 평균적으로는 종단적 계통 표집과 대표구 방법이 총화 대표구와 단순 계통 표집방식보다 우수해 보인다. 하지만 대표구의 경우, 분산이 가장 크며 특히 극단치가 많은 문제가 있다. 만약, 실제 과정에서 종단적 계통 표집을 사용할 수 없는 경우라면, 대표구 표집방법을 포기하고 단순 계통 표집이나 총화 대표구 방식을 사용하는 것이 더 적절할 수도 있는 것이다. 이럴 경우, 여러 가지 요인을 고려해 보아야 하겠지만, 투표소를 충분히 뽑을 수 있다면 단순 계통 표집방식이, 그렇지 못할 경우는 총화 대표구 방식이 더 효율적일 수 있다.

〈표 9〉 표집방법별 예측오차의 크기와 극단치 수

구분	단순 계통 (n=190)	대표구 (n=190)	총화 대표구 (n=190)	종단적 계통 (n=190)
평균	2.3	1.9	2.2	1.4
분산	4.1	7.9	5.1	2.2
극단치 사례 수	2개	9개	5개	2개

\* 극단치:  $q_3 + 3(q_3 - q_1)$ 보다 큰 값

지금까지의 연구결과를 종합해 보면, 다음과 같이 정리할 수 있다. 비록 표집방법별 예측오차의 평균이 통계적으로는 유의미하지 않았지만 실무적인 관점에서 보면 종단적 계통 표집과 대표구 표집의 예측오차 차이가 의미있는 것으로 받아들여진다. 그리고 극단치 분석결과까지 종합적으로 고려할 경우, 평균과 분산 두 가지 측면 모두에서 종단적 계통 표집이 4개 표집방법 중 가장 우수한 방법인 것으로 판단된

다.

이러한 연구결과는 지금까지의 투표소 추출방법이 적절치 않았다는 점을 보여 준다. 지금까지는 대표구 추출방법과 층화 대표구 방법 중에서 어느 방법이 효율적인가에 대한 논란이 많았고, 실제로 이 두 방법을 주로 적용해 왔다. 실제 분석결과는 이러한 가정이 타당하지 않았다는 것을 보여 준다.

본 연구에서 대표구 방법의 정확도가 종단적 계통 표집방법에 비해 높게 나타나지 않았는데, 이것은 선거구 전체의 투표결과를 언제나 나타내 주는 그러한 대표구가 존재하지 않는다는 투프테와 선(Tufte & Sun 1975)의 연구결과와도 일치한다. 즉 선거구는 끊임없이 변화되고 있고, 따라서 어떤 특정 투표소가 전체 선거구의 투표결과와 유사했다고 해서 그러한 특성이 다음 선거에서 계속된다고 기대할 수 없다는 것이다. 특히 본 연구의 결과가 투표소 조정이 거의 없었던 지역을 대상으로 실시한 분석에서 얻어진 것이라는 점을 감안하면 실제로 대표구 추출방법의 유용성은 상당히 낮다고 평가된다. 층으로 구분해서 대표구를 추출할 경우 대표구 추출방법의 많은 문제점이 해결될 것이라던 주장도 설득력이 떨어지는 것으로 나타났다.

또한 단순 계통 표집방법의 경우도 표본크기가 충분히 확보되지 않을 경우, 대표구 방법보다 예측력이 떨어질 것이라는 예상이 사실로 드러났다. 분석결과 단순 계통 표집방법을 사용할 경우, 13개 정도의 투표소를 뽑아야 대표구 방식으로 7~9개 정도의 투표소를 뽑을 경우와 비슷한 예측오차를 보였다.

### 3. 종단적 계통 추출방법과 층화

지금까지 분석해 온 바와 같이 투표소 추출방법으로는 그동안 실제 적용하지는 않았던 종단적 계통 표집방법이 가장 뛰어난 것으로 나타났다. 이 방법은 과거 선거결과를 활용하고는 있지만 대표구 방법과는

달리 다양한 성향의 투표소를 표집하고 있고, 또 층화 대표구와는 달리 층 구분이 임의적이지 않다는 장점이 있다.

이러한 연구결과에도 불구하고 각 표집방법별로 장단점이 존재한다는 점을 염두에 둘 필요가 있다. 결국 종단적 계통 표집방법은 층을 무시하는 데서, 층화 대표구 방식은 층 구분의 임의성이 개입될 수 있다는 점에서, 그리고 대표구 표집방식은 투표소 변화를 무시하는 데서 각각 예측오차가 커지게 된다. 그렇게 본다면 명확하게 층이 구분되는 경우에는 층화 방식이, 그렇지 않은 경우에는 종단적 방식이 적절한 선택이라고 볼 수 있겠다.

〈표 10〉 종단적 계통 표집방법 변화에 따른 예측오차 비교(철원/화천/양구/인제 선거구)

구 분	9개 추출	11개 추출	13개 추출	평균
층화 대표구	4.45	3.97	3.34	3.92
종단적 계통	9.58	0.84	6.67	5.70
층화후 종단적 계통*	0.39	2.87	2.06	1.77

\* 개별郡을 하나의 층으로 간주한 후 각 층에서 종단적 계통 표집방법으로 투표소를 추출하여 각 층(개별郡)의 값을 계산한 후 이를 합산하는 방식

〈표 10〉은 비교적 층이 명확하게 구분된다고 평가받고 있는 복합선거구를 대상으로 분석한 결과이다. 총선에서의 복합선거구는 대부분 소지역별로 자기 지역 출신 후보를 지지하는 등 명확히 대비되는 투표행태를 보이기 때문에 지역(郡) 구분과 층의 구분이 거의 일치하는 특징을 보인다. 한 개 선거구의 결과만을 가지고 일반화시킬 수는 없겠지만, 위의 분석결과는 층이 존재할 때의 표집상의 문제점을 보여 준다.

층화 대표구와 종단적 계통 표집방법은 앞의 분석에서 사용한 방식이며, 층화후 종단적 계통 표집은 이 두 가지 방법을 혼합한 방식이

다. <표 10>에서 살펴볼 수 있듯이 층이 명확하게 구분되는 선거구의 경우는 기존의 종단적 계통 표집방법보다는 층화 대표구 방식이 상대적으로 우수해 보인다. 하지만, 층화 표집의 장점과 종단적 계통 표집의 장점을 결합한 방식으로 추출할 경우에는 예측오차가 더 줄어드는 것으로 나타났다.

이와 같은 복합선거구가 아니더라도 층이 비교적 명확이 구분되는 선거구의 경우, 예를 들어 읍/면/동별로 지지패턴이 뚜렷히 나뉘는 경우는 유사 읍/면/동별로 층을 구분한 후 층화후 종단적 계통 표집방법을 사용하면 유용할 것으로 판단된다.

## V. 결론

본 연구에서 분석자료로 사용한 17대 총선이 투표소 선정과 관련하여 어느 정도 보편성을 띠고 있는지에 대해서 검토할 필요가 있다. 일단, 17대 총선의 경우 탄핵정국이라는 특수성이 국회의원 선거라는 보편성을 지배했던 것처럼 보인다. 이러한 특수성으로 인해 17대 총선결과를 토대로 실시한 본 연구의 결과는 판단표집방법에 불리하게 나타났던 것으로 생각할 수 있다. 이번 17대 총선의 분위기가 과거 선거와 달라진 만큼 과거 선거결과에 기반한 판단표집 방법은 오류가 커질 수 있기 때문이다. 하지만, 이러한 특수성은 지난 다른 선거에서도 폭풍, 병풍 등의 형태로 선거결과에 커다란 영향을 미쳐 왔다. 그리고 앞으로 실시될 미래의 선거에서도 또 다른 이슈가 선거에 영향을 미치게 될 것으로 보인다. 2008년 18대 총선 역시, 탄핵은 아니지만 또 다른 어떤 변화가 있을 수 있기 때문이다. 따라서 이번 17대 총선 자료가 특별히 판단표집 방법에 불리한 자료라고 판단할 필요는 없을 것으로 본다.

또 본 연구의 한계로 지적될 수 있는 것은 대통령 선거를 토대로



투표소를 추출하였다는 점이다. 동일 유형의 선거자료를 활용할 경우 좀더 대표구 방법에 의한 투표소 추출방법에 유리한 결과가 나왔을 가능성도 있다. 총선의 경우 후보자와의 관련성에 따라 투표행태가 영향을 많이 받을 수 있는데 이러한 투표소별 특성이 대통령 선거나 광역단체장 선거에서는 강하게 나타나지 않기 때문이다. 따라서 대통령 선거에서의 투표행태를 토대로 투표소별 특성을 분석할 경우 총선에만 나타나는 특성이 반영되지 않게 되고 따라서 투표소의 대표성이 달라질 가능성이 있다.

그러나 한국의 경우 투표소 조정이 계속되고 있기 때문에 비단 이번 17대뿐만 아니라 과거 16대의 경우도 직전의 동일 유형의 선거자료를 활용하는 것에 어려움이 많다. 동일 종류의 선거자료를 토대로 대표구 추출을 하기는 어렵기 때문에 본 연구의 결과가 대표구 추출방법에 실제보다 불리하게 설계되었다고 보기는 어렵다고 본다. 따라서 비록 대통령 선거자료를 이용해서 판단표집을 실시했고, 또 17대 총선이 탄핵정국이라는 특수성이 있지만 본 연구의 결과는 어느 정도의 일반성을 가진다고 볼 수 있다.

이러한 점을 고려하여 본 연구결과를 토대로 출구조사의 투표소 추출방법에 대해 결론을 내려보면 다음과 같다. 앞으로의 출구조사는 종단적 계통 표집방법을 사용하는 것이 합리적이라고 볼 수 있다. 과거에 가장 보편적으로 사용되어 왔던 대표구 방식은 앞으로 사회변화가 더욱 가속화될 것이고 투표행태의 변화도 더 많을 것으로 예상되기 때문에 그 효율성이 떨어질 것으로 예상된다. 종단적 계통 표집방법은 우리나라에서는 아직 적용되지 않았지만 이미 영국에서 활용하고 있는 방법이다. 다만 앞서 지적하였듯이 상이한 특성의 층이 이미 존재하고 있을 때 이러한 존재를 무시하고 계통적으로만 추출할 경우 특정 층이 과대 대표될 가능성이 있다. 예를 들어 특정 선거구가 비슷한 크기의 두개 층으로 구분되어 있는데 여기에서 다섯 개의 투표소를 추출할 경

우 어느 한 층이 다른 층보다 더 많이 대표되게 된다. 따라서 실제 상이한 유형이 존재할 경우 각 유형별로 구분하여 추출하는 것이 바람직하다. 이런 경우, 단순 평균 형식의 추정방법을 사용하면 편향이 생기게 되므로 층별 추출률을 반영한 층화추출에서의 추정식을 사용해야 한다.

그러나 층화 대표구 추출방법에 대한 검토결과에서 보았듯이 무조건 층을 구분하는 것은 오히려 예측오차를 증가시킬 수 있다. 즉 실제로 층이 존재하지 않을 경우 또는 존재하더라도 그 수나 크기를 잘못 추정하는 경우에는 예측오차가 커지기 때문이다. 따라서 층이 존재한다는 명확한 경우를 제외하고는 층을 고려하지 않는 것이 바람직할 것으로 보인다. 출구조사의 표본크기에 대해서는 많은 논란이 있었지만 명확한 결론이 내려지지 못한 상태였다. 이번 연구는 그러나 지금까지 사용해 왔던 7개에서 10개의 투표소 수가 가장 타당했었다는 것을 보여 준다. 이론적으로 본다면 투표소 수가 증가할 수록 표집오차는 감소한다고 보아야 한다. 그러나 본 연구의 결과는 9개를 초과해서 추출했을 때 표집오차가 감소하지 않았다. 표집방법과는 달리 표본의 크기에서는 그동안의 가정이 적절했었다는 것을 보여 주는 결과다.

하지만, 이러한 것 모두가 실제 선거예측 과정에서 100% 적용 가능한 것은 아니다. 즉, 종단적 계통 표집방법으로 9개 투표소를 추출하는 것이 최선의 방법만은 아니라는 것이다. 먼저, 선거의 성격, 투표소 변화 정도, 소요예산, 조사회사의 실사능력 등을 종합적으로 고려해야 한다. 다음으로 표집방법별 장단점을 이러한 전제조건과 대비시킨 후 선거구별로 적절한 표집방법을 선택해야 한다. 이때 고려해야 할 사항은 각 선거구별 예상 경합도에 따른 오차의 허용 범위이다. 실제상황에서는 투표소 표집오차 이외에 투표소 내에서 응답자 추출과정에서 나타나는 표본추출오차, 그리고 실사에서 발생하는 무응답 등에 따른 비표본오차도 발생한다는 점을 함께 고려해야 한다. 마지막으로

선택된 표집방법에 잘 부합되는 방향으로 추출 투표소 수를 결정하면 될 것이다. 앞으로 투표소 추출에 따른 예측오차를 감소시키기 위해서는 보다 다양한 선거와 방법론에 대한 실증적 연구 검토가 이루어질 필요가 있다.

## 참고문헌

- Biemer, Paul, R. Folsom, R. Kulka, J. Lessler, B. Shah, and M. Weeks. 2003. "An Evaluation of Procedures and Operations Used by the Voter News Service for the 2000 Presidential Election," *Public Opinion Quarterly*, 67(1):32-45.
- Moon, Nick. 1999. *Opinion Polls: History, Theory and Practice*. Chapter 7. Manchester University Press.
- Tufte, E. R. and A. Sun. 1975. Are There Bellwether Electoral Districts? *Public Opinion Quarterly*, 39(1):1-18.
- 김정훈. 2003. "선거예측과 출구조사: 16대 대선을 중심으로." 《조사연구》 4(2): 87-102.
- 김재광. 2004. "제17대 총선 예측조사와 관련된 통계적 이슈 고찰." 《조사연구학회 춘계학술대회 발표논문집》 203-207.
- 류제복. 2003. "출구조사의 역사와 개선방향." 《조사연구》 4(1):31-48.
- 조성겸. 2001. "2천년 총선 예측조사와 편향요인," 《한국언론학보》 45(3): 328-359.
- 한상태·강현철·허명희. 2004. SPSS 실험설계와 분산분석. SPSS아카데미.
- 홍내리·허명희. 2001. 제16대 국회의원 선거의 예측조사에 대한 사후적 검증. 조사연구 2(1):1-36.