

연구논문

## 출구조사를 이용한 제21대 국회의원 선거 의석 수 예측결과 분석\*

천승호\*\* · 임요한\*\*\* · 박민규\*\*\*\*

2020년 4월에 실시한 제21대 국회의원 선거에서 출구조사 결과를 이용한 각 정당의 의석 수 예측은 크게 빛나갔다. 본 연구에서는 출구조사를 이용한 의석 수 예측의 문제점을 분석하였다. 이를 위해, 제21대 국회의원 선거결과를 다양한 관점에서 분석하였고, 최근 3번의 국회의원 선거양상의 변화를 통해 21대 총선과 이전 국회의원 선거의 차이와 공통점을 살펴보았다. 제21대 국회의원 선거는 사전투표제도가 도입된 이후 시행된 선거 중에서 가장 사전투표율이 높은 선거였고, 결과적으로 높은 사전투표의 비중으로 인해 출구조사 결과만을 이용한 예측에 체계적인 편향이 발생하게 되었다. 즉, 제21대 국회의원 선거의 출구조사는 실제 투표자 전체에 대한 대표성을 갖추지 못한 일종의 추출틀 오차로 인해 예측에 체계적인 오류가 발생한 것으로 확인하였다.

주제어: 국회의원 선거, 출구조사, 체계적 편향, 베이지 추론, 사전투표

\* 본 연구는 천승호의 석사학위논문 내용을 바탕으로 이루어짐.

\*\* 고려대학교 통계학과 석사과정(cheon2456@korea.ac.kr), 제1저자.

\*\*\* 고려대학교 통계학과 석박사통합과정(lyh2744@korea.ac.kr), 제2저자.

\*\*\*\* 고려대학교 통계학과 교수(mpark2@korea.ac.kr), 교신저자.

## I. 서론

출구조사(exit poll)는 선거 당일 투표를 마친 유권자의 일부를 확률표본추출(probability random sampling)하여 투표 행위에 대한 조사를 수행하는 선거 여론조사 방법이다. 국내에서는 2000년 제16대 국회의원 선거에서 최초로 전국 단위의 출구조사가 이루어졌고, 2010년 제5회 전국동시지방선거부터 지상파 방송 3사(KBS, MBC, SBS)와 방송협회가 구성한 방송사 공동예측조사위원회(KEP: Korea Election Pool)를 통해 출구조사를 수행해 왔다. 출구조사를 위해서는 확률표본추출 방법을 통해 각 선거구에서 투표소를 추출하고, 추출된 투표소에서 투표한 유권자 중 일부를 추출해 조사한다. 제21대 국회의원 선거의 방송사 공동 출구조사에서는 투표소를 마치고 나오는 개인 중 매 5번째 투표자를 같은 간격으로 조사하는 계통표본추출(systematic sampling) 방식으로 표본을 추출해 조사를 수행하였다.

선거일 이전에 유권자를 모집단으로 정의하여 수행되는 여론조사와 달리, 선거 당일 투표자를 모집단으로 정의하여 수행되는 출구조사는 가장 신뢰할 수 있는 선거 예측조사라고 할 수 있다. 하지만 대통령 선거나 전국동시지방선거와는 달리 출구조사를 통한 선거결과 예측의 정확성이 낮은 국회의원 선거에서는 출구조사의 효율성에 대한 의심이 제기되어 왔으며, 제21대 국회의원 선거에서도 출구조사를 통한 정확한 선거결과 예측에 실패하며 비판을 받게 되었다.

출구조사 결과를 이용한 선거결과 예측은 크게 두 단계에 걸쳐 이루어진다. 1단계에서는 추출된 투표소로부터 투표를 마치고 나오는 투표자 중 일부를 확률표본설계를 통하여 선택하여 조사자료를 수집한다. 2단계에서는 수집된 자료를 이용하여 선거결과를 예측하는 단계로서, 무응답과 사전투표를 포함한 선거의 다양한 요인들을 고려하여 국회의원 선거의 경우에는 정당의 의석 수를 예측하게 된다. 즉, 2단계에서는 1단계에서 이루어진 조사 혹은 수집된 자료가 가질 수 있는 편향을 조정하는 과정으로 이해할 수 있다.

두 번째 단계에서 이루어지는 출구조사를 통한 선거결과 예측에 관련된 기존 연구에서는 베이즈(Bayes) 추론방법을 통해 당선가능도(chance of winning)를 계산하고, 이를 이용한 정당 의석 수 예측방안이 주로 고려되었다. 이러한 방법은 의석 수

예측을 위하여 각 선거구별로 예측 지지도가 높은 정당에 속한 후보자의 수를 세는 단순 셈 방안보다 분포 생성을 통한 의석 수의 구간 추정이 가능하다는 장점이 있다. 허명희(2008)는 국회의원 선거에서 주요 정당 의석 수 예측을 위해 열세자 침묵을 보정하는 베イズ 추론 방법을 제시하였다. 또한 박민규·윤상영(2019)은 지역별 특성을 고려하여 과거의 선거결과에서 나타난 체계적 편향을 조정하는 과정을 추가하였다. 하지만 기존의 연구들은 과거의 선거에서 나타난 편향을 현재의 선거결과 예측을 위하여 적용하는 방법이기 때문에 선거 환경이나 정치 환경의 변화에 따라 그 사용은 제한적일 수밖에 없다.

본 연구에서는 제21대 국회의원 선거결과를 다양한 관점에서 살펴보고 출구조사를 이용한 정당 의석 수 예측의 문제점을 분석하였다. 특별히 제21대 국회의원 선거와 이전 국회의원 선거에서 나타난 투표행위의 근본적인 차이를 살펴보고 이를 기반으로 출구조사를 통한 선거결과 예측 실패의 원인을 분석하였다. 본 연구를 통해 산출된 분석의 결과로 출구조사를 이용한 의석 수 예측을 위한 또 다른 방법을 제안할 수 있을 것이다.

## II. 국회의원 선거 분석<sup>1)</sup>

### 1. 선거결과 분석

#### 1) 제21대 국회의원 선거결과

2020년 4월에 실시된 제21대 국회의원 선거의 결과와 출구조사를 이용한 KBS<sup>2)</sup>의 예측결과는 <표 1>과 같다. 지역구 선거에서는 21개 정당이 후보를 내었고, 비례대표 선거에서는 35개 정당이 출마하였다. 이 중 1명 이상의 당선인을 보유한 정당만을 표에 표시하였다. 또한, 더불어민주당(지역구)과 더불어시민당(비례대표), 미래통합당(지역구)과 미래한국당(비례대표)은 당선인 수를 합산하여 계산하였다.

1) 본 장의 내용은 중앙선거관리위원회 선거통계시스템에 등록된 자료를 분석한 결과임.

2) 지상파 3사가 모두 결과를 방송했으나 결과가 유사하여 본 연구에서는 KBS 결과만을 활용함.

<표 1> 제21대 국회의원 선거결과와 예측 결과 비교(단위: 명)

정당	당선인 수	예측결과
더불어민주당 + 더불어민주당	180	155~178
미래통합당+미래한국당	103	107~130
정의당	6	5~7
국민의당	3	2~4
열린민주당	3	1~3
무소속	5	1~7

<표 2> 21대 총선 지역구 선거 시·도별 당선자 수 (단위: 명)

시·도	더불어 민주당	미래 통합당	정의당	무소속	합계
강원도	3	4		1	8
경기도	51	7	1		59
경상남도	3	12		1	16
경상북도		13			13
광주광역시	8				8
대구광역시		11		1	12
대전광역시	7				7
부산광역시	3	15			18
서울특별시	41	8			49
세종특별자치시	2				2
울산광역시	1	5			6
인천광역시	11	1		1	13
전라남도	10				10
전라북도	9			1	10
제주특별자치도	3				3
충청남도	6	5			11
충청북도	5	3			8
합계	163	84	1	5	253

출구조사를 통해 여당의 과반 압승과 원내 제1당 예측은 성공하였지만, 선거 전 여론조사에서부터 두 정당의 지지율 차이가 확연히 나타난 것을 고려하면 1당 예측 성공에 큰 의미는 없는 것으로 보인다. 그리고 정의당을 비롯하여 국민의당, 열린민주당, 무소속은 정확한 의석 수 예측에 성공하였지만, 실제 당선인 수가 많은 두 주요정당의 예측에 실패하여 출구조사 기반 예측이 성공적이라고 판단하기는 어렵다. 21대 총선 지역구 선거에서 17개 시·도별 각 정당의 당선자 수 분포는 <표 2>와 같다. 전체 253개 지역 선거구에서 더불어민주당이 163석을 차지하였고, 미래통합당이 84석을 차지하였다.

<표 3>은 사전투표제도가 도입된 이후의 국회의원 선거, 대통령 선거, 지방선거에서 사전투표율과 전체 투표율을 보여주고 있다. 또한, 선거의 종류에 따라 전체 투표율이 다른 경향이 있기 때문에 단순히 사전투표율의 수치만으로 비교하기는 무리가 있어, 전체 투표 중 사전투표가 차지하는 비중을 같이 구하였다.

<표 3> 역대 선거 투표율 비교

(단위: %)

선거 구분	사전투표율	전체 투표율	사전투표 비중
6회 동시지방선거(2014.06.04)	11.5	56.8	20.2
20대 국회의원선거(2016.04.13)	12.2	58.0	21.0
19대 대통령선거(2017.05.09)	26.1	77.2	33.8
7회 동시지방선거(2018.06.13)	20.1	60.2	33.4
21대 국회의원선거(2020.04.15)	26.7	66.2	40.3

이번 총선은 전체 66.2%의 투표율과 26.7%의 사전투표율을 기록했다. 일반적으로 총선이 대선에 비해 전체 투표율이 낮음에도 불구하고 이번 총선은 사전투표제도가 도입된 이후 시행한 모든 선거 중 가장 사전투표율이 높은 선거였고, 사전투표율이 차지하는 비중도 40%가 넘는 첫 선거해였다. 이런 현상을 보다 상세히 살펴보기 위하여 21대 총선 지역구 선거에서 17개 시·도별 투표율 결과를 살펴보고 이는 <표 4>와 같다.

서울특별시와 경상남도, 강원도는 사전투표율, 당일투표율, 전체 투표율이 모두 평균 이상이었으며, 인천광역시, 제주특별자치도, 충청남도는 투표유형별 모든 투표율이 평균 이하로 나타났다.

&lt;표 4&gt; 21대 총선 지역구 선거 시·도별 투표율

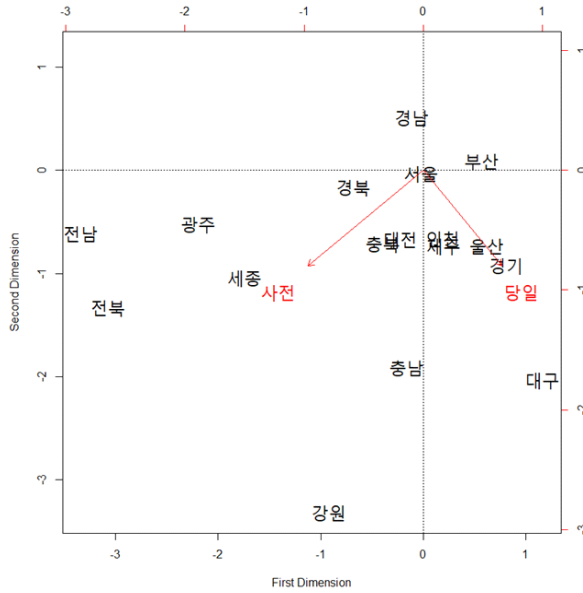
(단위: %)

시·도	사전투표율	당일투표율	전체 투표율
강원도	28.7	37.3	66.0
경기도	23.9	41.1	65.0
경상남도	27.6	40.2	67.8
경상북도	28.7	37.7	66.4
광주광역시	32.2	33.7	65.9
대구광역시	23.6	43.4	67.0
대전광역시	26.9	38.6	65.5
부산광역시	25.5	42.2	67.7
서울특별시	27.3	40.8	68.1
세종특별자치시	32.4	36.1	68.5
울산광역시	26.0	42.7	68.6
인천광역시	24.7	38.5	63.2
전라남도	35.8	32.0	67.8
전라북도	34.7	32.3	67.0
제주특별자치도	24.7	38.2	62.9
충청남도	25.3	37.1	62.4
충청북도	26.7	37.2	64.0
전국 평균	26.7	39.5	66.2

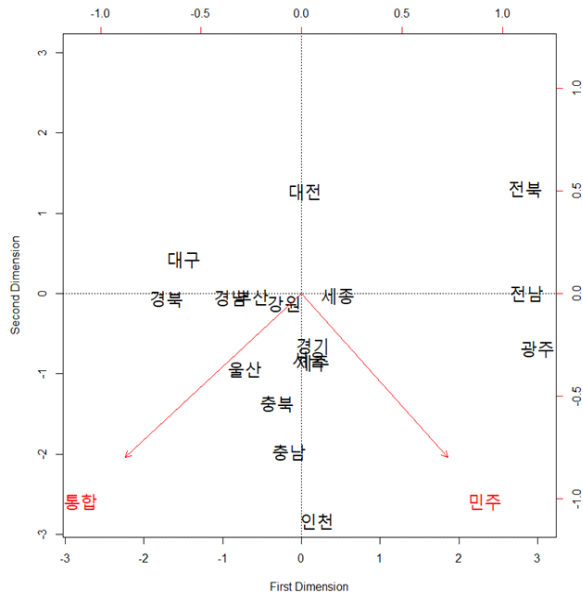
지역과 선호하는 투표유형, 지역과 선호하는 정당의 관계를 살펴보기 위해 대칭형 대응분석(symmetric correspondence analysis)을 수행하였다. 대칭형 대응분석은 행과 열이 대칭적인 경우로 행 변수와 열 변수 간 상관의 최대화에 분석의 목표를 두게 되고, 그 결과로 행 범주와 열 범주에 다차원 수치가 부여된다(허명희 2014).

<그림 1>에는 투표유형(사전투표, 당일투표)에 따른 17개 시·도별 투표수를 이용하였고, <그림 2>에는 더불어민주당, 미래통합당 두 정당의 17개 시·도별 득표수를 이용하였다. <그림 1>을 보면 지역별로 선호하는 투표유형을 알 수 있다. 전라남도, 전라북도, 광주광역시, 세종특별자치시는 다른 시·도와 비교하여 당일투표보다 사전투표를 선호하는 경향을 띠었다. 그에 반해 대구광역시, 경기도, 울산광역시, 부산광역시는 사전투표보다 당일투표를 선호하였다. 서울특별시와 경상남도는 투표유형별 모든 투표율이 높기 때문에 어느 한쪽에 치우치지 않고 그래프의 중앙에 배치되었고, 인천광역시, 제주특별자치도, 충청남도도 마찬가지로 투표유형별 모

든 투표율이 낮기 때문에 중앙에 배치되었다.



<그림 1> 21대 총선 지역과 투표유형의 대칭형 대응분석



<그림 2> 21대 총선 지역과 정당의 대칭형 대응분석

<그림 2>를 보면 17개 시·도별 정당 선호도를 알 수 있다. 대구광역시, 경상북도, 경상남도, 부산광역시, 울산광역시는 더불어민주당보다 미래통합당을 선호하는 경향을 보였고, 전라북도, 전라남도, 광주광역시는 미래통합당보다 더불어민주당을 선호하였다. 그리고 세종특별자치시, 경기도, 서울특별시, 제주도, 인천광역시, 제주도는 더불어민주당이 대다수 당선되었지만, 그래프의 중앙에 가까운 이유는 호남 지방(광주광역시, 전라남도, 전라북도)에 비해 득표수 차이가 작게 나며 경합을 벌였기 때문이다.

또한 <그림 2>를 통해 단순히 각 지역의 당선자 정당 비율을 보는 것이 아닌 해당 지역의 정당 성향 집중도를 알 수 있다. 예를 들어 세종의 경우 2개의 선거구 모두 더불어민주당 후보자가 당선되었다. 하지만 호남 지방에 비해 중앙에 위치한 이유는 득표 수를 기준으로 대칭형 대응분석을 했기 때문이다. 호남 지방에 비해 1, 2 위 후보자 간의 득표 수 차이가 작아 그래프의 중앙에 위치한 것이고, 이런 현상은 더불어민주당의 당선 비율이 높은 대전, 경기, 서울, 제주 등에서도 확인할 수 있다.

한 가지 눈여겨볼 수 있는 것은 <그림 1>과 <그림 2>에서 두 요인인 사전투표 여부 변수와 일반적으로 알려진 호남과 영남으로 구분이 되는 지역 변수를 이용한 대응분석 결과가 서로 유사하다는 것이다. 이는 사전투표 여부 변수와 지역 변수가 선거결과에 미치는 영향이 21대 국회의원 선거에서는 일정 수준 중첩(confounding)되어 나타나고 있다는 것이며, 이에 대한 상세한 내용은 다음 장에서 다루기로 한다.

## 2) 제19, 20 그리고 21대 국회의원 선거양상 변화

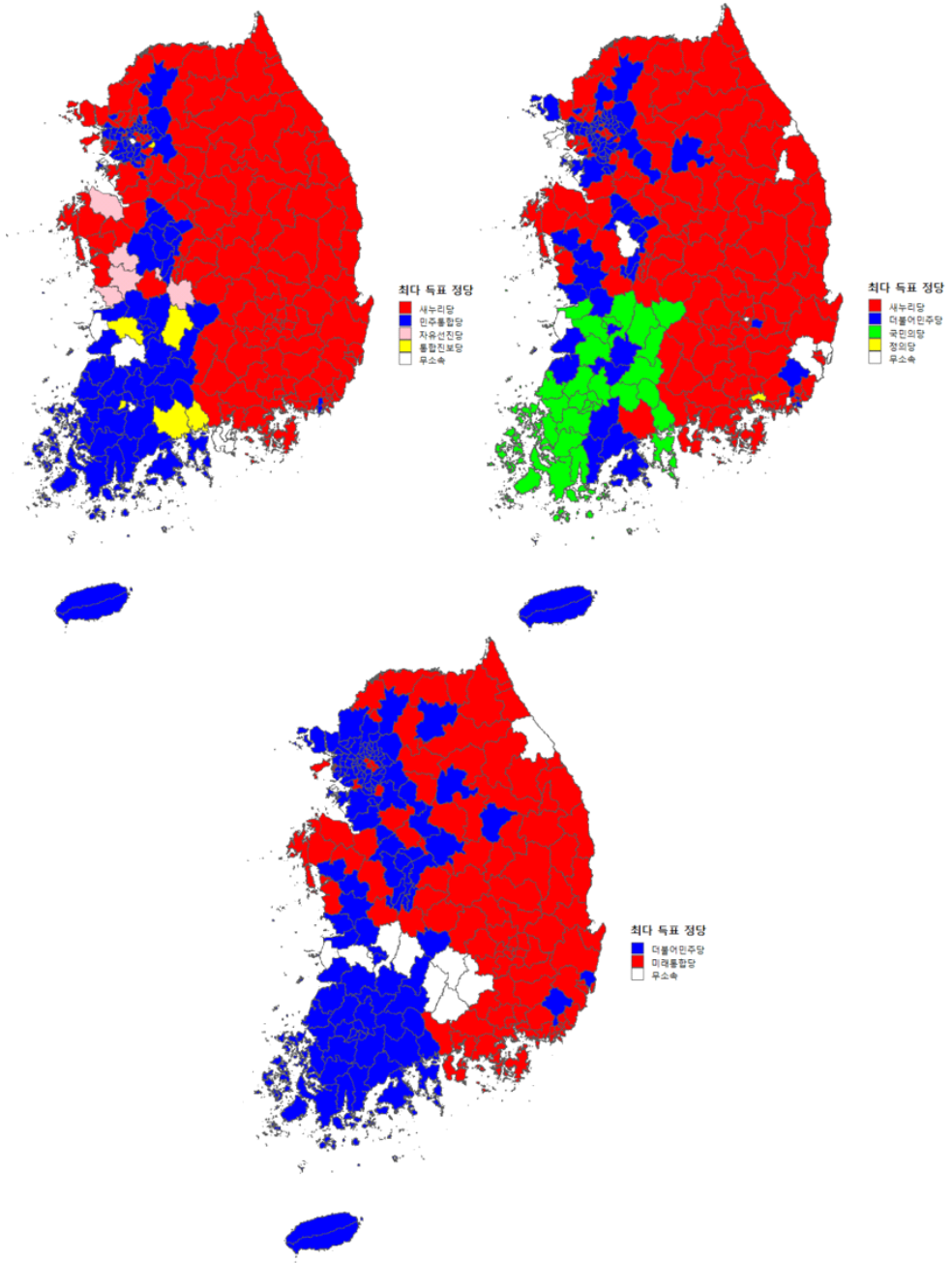
제21대 국회의원 선거와 과거의 선거 구도를 비교하기 위해 최근 3번의 국회의원 선거양상을 살펴보았다. <그림 3>은 19, 20, 21대 총선 지역구 선거결과를 전국 지도에 시각화한 것이다. 첫 행의 왼쪽이 19대, 오른쪽이 20대, 아래가 21대 선거결과이다. 이를 위해 대한민국 행정구역 시·군·구 SHP(shape file) 자료를 이용하였다. 시·군·구는 선거마다 크게 변하는 선거구에 비해 변동이 적고, 지도를 통해 시각적으로 파악하기 쉬운 장점이 있어 이를 사용하였다. 하지만 지역구 국회의원 선거는 시·군·구가 아닌 선거구를 기준으로 한다. 선거구는 지리적 경계를 바탕으로 인구 상하한선을 설정하고, 선거구를 분구·합구·조정하여 확정된다. 따라서 한 시·군·구가 두 개 이상의 선거구로 나뉠 수도 있고, 여러 시·군·구가 하나의 선거구 안에 포함될 수도 있다.



전체적인 경향을 보기 위한 것이므로 편의를 위해 지도에 표시하기 힘든 경상북도 울릉군과 인천광역시 옹진군을 제외한 248개 시·군·구를 사용하였다. 각 시·군·구에 선거구가 대응되면 이들의 득표 수를 합해 최다 득표 정당을 구하였고 따라서, 지역구 선거결과와 조금 차이가 있을 수 있다. 예를 들어, 21대 총선에서 경기도 고양시 갑 선거구에 정의당 후보가 당선되었지만, 고양시 덕양구는 경기도 고양시 갑 선거구와 경기도 고양시 을 선거구가 합쳐졌기 때문에 두 선거구 득표 수를 합산하여 최다 득표 정당에 정의당이 선정되지 않았다.

선거구 결과를 대응시켜 최다 득표 정당을 구하였고, 한 시·군·구에 두 개 이상의 그림을 볼 때 주의할 점은 선거구와 시·군·구를 혼동하지 않아야 한다는 것이다. 21대 총선의 지역구 선거에서 더불어민주당은 163석, 미래통합당은 84석을 차지했지만 그림으로 보면 오히려 미래통합당이 더 많은 의석을 차지한 것처럼 보인다. 이는 더불어민주당이 대부분의 의석을 차지한 경기도와 서울의 선거구 수는 각각 59개, 49개로 전체 253개 선거구의 40% 이상을 차지하지만, 인구밀집도가 높고 선거구 면적이 작아 지도상에 작게 나타나기 때문이다.

19대에서 20대, 20대에서 21대로 올수록 보수 정당의 당선이 감소하고, 민주당계 정당의 당선이 증가하는 추세를 확인할 수 있다. 특히 수도권인 서울특별시, 경기도, 인천광역시는 시간 흐름에 따라 점차 민주당계 정당 지지가 증가하는 추세가 확연히 나타난다. 최근 3번의 총선에서 영남 지방(경상남도, 경상북도, 대구광역시, 부산광역시, 울산광역시)은 보수 정당을 선호하는 경향이 뚜렷하게 나타나며, 호남 지방(전라남도, 전라북도, 광주광역시)과 제주도는 민주당계 정당을 선호하는 경향이 뚜렷하게 나타난다. 그러나 충청남도와 충청북도는 어느 한쪽으로 선호 정당이 물리는 현상이 3번의 선거 동안 나타나지 않았다. 이번 총선만의 특징으로 대전과 세종에서 압도적인 민주당계 정당 선호 현상이 나타났고, 이는 19대와 20대 총선에서는 볼 수 없었던 현상이다.



<그림 3> 19, 20, 21대 총선 시·군·구별 최다 득표 정당 비교



<그림 4> 19, 20, 21대 총선 경합지역 비교

<그림 4>는 19, 20, 21대 총선 지역구 선거결과에서 각 시·군·구의 1, 2위 정당 득표율 차이를 이용하여 경합 지역을 나타낸 그림이다. 첫 행의 왼쪽이 19대, 오른쪽이 20대, 아래가 21대 결과이다. 득표율 차이 구분은 1%p, 3%p, 5%p, 10%p를 경계점으로 하였고, 색이 진할수록 득표율의 차이가 작아 박빙이었음을 의미한다.

그림을 보면 세 번의 선거에서 20대 총선이 가장 경합이 심했음을 알 수 있다. 실제로 당시 더불어민주당이 123석, 새누리당이 122석을 차지했다. 20대 총선만큼은 아니지만 21대 총선에서 수도권과 충청남도, 충청북도, 강원도에서 경합이 두드러졌다. 새누리당이 152석, 민주통합당이 127석을 차지한 19대 총선보다 의석 수가 확연히 많은 차이가 났음에도 불구하고 여러 시·군·구에서 경합 양상이 벌어진 것을 확인할 수 있다. 이는 선거 예측을 더 어렵게 한 요인 중 하나로 볼 수 있다. 즉, 지난 세 번의 국회의원 선거결과를 비교해 볼 때, 지역 특성과 경합 지역의 분포는 정도의 차이는 있으나 큰 변화가 없는 것으로 나타나고 있다. 즉, 21대 국회의원 선거 시 수행된 출구조사 예측 오차가 근본적으로 투표 행위의 변화에 기인한 것으로 판단하기는 그 근거가 부족한 것으로 판단된다.

## 2. 출구조사를 기반으로 한 21대 총선 의석 수 예측의 한계

국회의원 선거에서 출구조사 결과를 기반으로 한 의석 수 예측이 실패하는 원인은 다양하다. 대통령 선거나 지방선거와 달리 당선자를 예측해야 하는 선거구가 많은 것이나 선거구 1곳당 출구조사를 위한 표본 수가 상대적으로 적은 것 등 국회의원 선거 자체가 갖는 한계가 있다. 이로 인해 득표율 차이가 적은 경합 지역의 예측이 어려워진다. 이번 총선에서는 신종 코로나바이러스 감염증(COVID-19)으로 인한 선거환경 변화에 따른 영향도 배제할 수 없다. 또한, 출구조사에서는 조사 거절로 인해 체계적인 편향이 발생하는 것이 예측 실패의 주요한 원인일 수 있다(이준웅 2004). 21대 총선에서도 COVID-19 감염을 우려한 응답 거절로 인해 체계적인 편향이 발생했을 가능성 역시 배제할 수 없다. 이를 살펴보기 위하여 19, 20 그리고 21대 국회의원 선거 시 수행한 출구조사의 응답 거절율을 살펴보고, 이는 각각 19대 21.7%, 20대 22.0% 그리고 21대 24.6%로 나타났다. 전체적으로 거절율이 증가하고는 있으나 COVID-19 상황을 고려할 때 그 차이가 유의한 것으로 판단하기는 어렵다. 무응답 조정을 통한 예측 방안의 타당성에 대해서도 본 장에서 살펴보겠다.

본 논문에서는 제21대 국회의원 선거결과 예측 실패의 원인 분석을 위해서 출구

조사 결과를 이용하여 산출된 추정량을 중점적으로 살펴보았다. 즉, 전통적인 방법인 성별 그리고 연령그룹 단위의 조정이 출구조사 결과에 적용되어 산출된 의석 수 예측결과에 대한 분석을 수행하였다. 이러한 분석을 통해서 전통적인 출구조사와 그 보정 방안을 통해 산출된 의석 수 추정량의 편향을 이번 선거에서 나타난 특이한 현상인 높은 사전투표 비중과의 연관성에서 찾고자 하였다.

<표 5> 21대 총선 출구조사 결과와 실제 결과 당선자 수 비교 (단위: 명)

출구조사	실제 결과			
	더불어민주당	미래통합당	정의당	무소속
더불어민주당	152	2	0	0
미래통합당	11	82	0	1
정의당	0	0	1	0
무소속	0	0	0	3

출구조사 결과를 이용한 예측과 실제 선거결과를 <표 5>에서 비교했다. 표에서 언급한 출구조사 결과는 조사를 통해 산출한 득표율이 1위인 후보의 소속 정당을 기준으로 작성되었으며, 예를 들어 <표 5>의 더불어민주당에 대응하는 행은 출구조사에서 더불어민주당의 당선을 예측한 총 154개 선거구 중 실제 152개 선거구에서 더불어민주당이 당선되었고, 나머지 2개 선거구는 미래통합당이 당선된 것을 나타낸다. 예측에 실패한 총 14곳의 선거구 중 11곳의 선거구는 미래통합당 당선을 예측했지만, 더불어민주당이 당선되었다. 출구조사 결과로 볼 때 더불어민주당 후보의 당선이 예측되었으나 미래통합당 후보가 당선된 곳이 2곳임을 고려하면 출구조사 결과를 이용한 예측에 체계적인 편향이 있었음을 알 수 있다. 예측에 실패한 14곳의 선거구에 대한 상세한 정보는 <표 6>과 같다.

<표 6>을 살펴보면 예측에 실패한 선거구는 어느 특정 지역에 몰려있기보다는 전국적으로 분포한 것을 확인할 수 있다. 다만, 정치성향이 가장 뚜렷한 호남 지방과 대구를 제외한 영남 지방에서 예측에 실패한 선거구는 없었다. 초경합 선거구라고 할 수 있는 득표율 차이 1%p 이하의 선거구는 14곳 중 단 한 곳뿐이었고, 서울 영등포구를 선거구에선 5%p를 넘겨 예측이 크게 부정확했음을 알 수 있다. 대부분 초경합 선거구가 아닌 곳에서 예측이 실패한 것과, 실패한 14곳 중 11곳이 미래통합당 후보의 당선을 예측했지만 더불어민주당 후보가 당선된 사실을 통해 출구조사

결과를 이용한 예측에는 체계적인 편향이 있음을 유추할 수 있다. 다만 이러한 체계적인 편향의 원인이 과거와 같이 무응답이나 응답자로부터 발생한 측정오차에 기인한 것인지에 대한 검토는 필요하다.

<표 6> 21대 총선에서 예측에 실패한 선거구 결과

선거구	출구조사 결과	실제 결과	1, 2위 득표율 차이 (%p)
경기 성남시 분당구 을	미래통합당	더불어민주당	2.84
경기 평택시 갑	미래통합당	더불어민주당	2.81
경남 양산시 을	미래통합당	더불어민주당	1.68
대구 수성구 을	미래통합당	무소속	2.74
대전 대덕구	미래통합당	더불어민주당	3.15
대전 동구	미래통합당	더불어민주당	3.45
대전 중구	미래통합당	더불어민주당	2.13
서울 영등포구 을	미래통합당	더불어민주당	5.91
서울 용산구	더불어민주당	미래통합당	0.66
서울 중구·성동구 을	미래통합당	더불어민주당	4.69
인천 연수구 을	미래통합당	더불어민주당	2.29
충남 논산시·계룡시·금산군	미래통합당	더불어민주당	4.67
충남 보령시·서천군	더불어민주당	미래통합당	1.73
충북 청주시 서원구	미래통합당	더불어민주당	3.07

이를 살펴보기 위하여 <표 7>에서는 2020 총선 출구조사 평가회의 자료에서 산출된 최종 당선자의 소속 정당 예측에 실패한 선거구의 투표 당일 및 사전투표 후보자 득표율 예측값과 실제 결과를 살펴보았다. <표 7>의 사전투표 출구 예측은 당일 출구조사 결과와 사전투표자의 성과 연령 분포를 이용하여 예측된 값이며, 당일 투표 예측은 당일 투표자의 성과 연령 분포를 이용한 무응답 조정, 5시 이후 투표자 보정이 이루어진 예측값을 나타낸다. <표 7>의 민주는 더불어민주당을, 통합은 미래통합당을 각각 나타내며, 대구 수성구 을의 경우에는 미래통합당 후보가 아닌 무소속 후보의 지지율이 기록되어 있다.

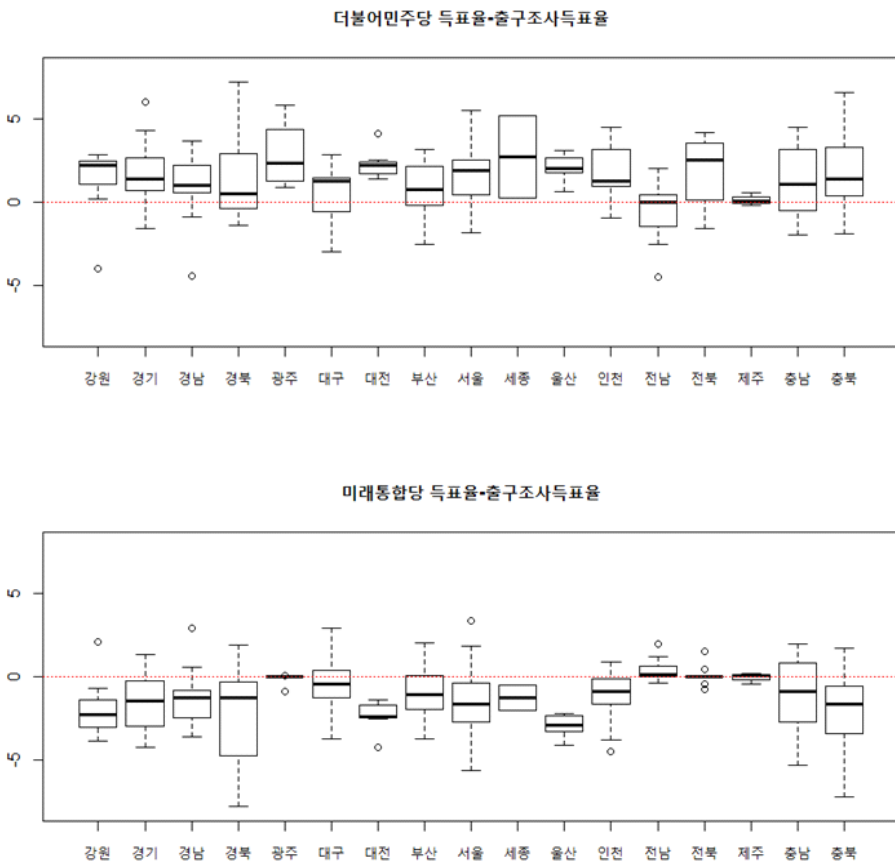
&lt;표 7&gt; 예측실패 선거구 사전투표 및 당일투표 예측결과 (단위: %)

선거구	당일투표				사전투표			
	출구 예측		실제 결과		출구 예측		실제 결과	
	민주	통합	민주	통합	민주	통합	민주	통합
경기 성남시 분당구 을	44.4	49.4	41.9	50.4	45.8	48.0	57.2	36.6
경기 평택시 갑	48.9	49.1	47.2	50.4	48.8	49.2	55.8	42.0
경남 양산시 을	48.4	48.2	45.0	51.0	47.7	48.5	56.0	40.5
대구 수성구 을	39.9	36.3	37.8	35.5	38.7	36.4	32.8	40.5
대전 대덕구	48.4	47.7	45.1	50.3	48.0	47.9	55.9	40.2
대전 동구	48.1	50.7	46.0	52.4	49.7	49.0	57.6	41.2
대전 중구	47.7	51.0	45.4	52.9	50.7	48.0	57.1	41.6
서울 영등포구 을	47.3	47.9	44.1	50.1	46.3	48.9	58.7	36.4
서울 용산구	45.8	48.0	40.8	53.8	48.8	45.4	55.4	40.0
서울 중구·성동구 을	49.1	50.4	46.3	52.9	50.1	49.4	59.7	39.6
인천 연수구 을	38.4	40.1	37.9	42.1	39.9	39.7	48.5	34.9
충남 논산시·계룡시·금산군	45.6	53.0	45.6	51.7	48.9	49.8	57.9	39.5
충남 보령시·서천군	50.7	49.3	44.8	55.2	51.7	48.3	53.5	46.5
충북 청주시 서원구	47.0	50.1	45.3	51.1	48.7	48.3	56.6	40.4

출구조사를 기반으로 하여 기존 방법을 통해 예측된 당일투표 예측값은 실제 결과와 매우 유사함을 알 수 있다. 당일 기준 지지율 1위를 예측하지 못한 곳은 두 곳임을 알 수 있다. 그러나 사전투표 결과의 예측값은 예측오차가 매우 크게 나타나고 있음을 확인할 수 있다. 이는 유의한 수준의 사전투표 득표율 예측오차가 높은 사전투표율에 반영되어 최종 출구조사 기반 예측실패에 영향을 준 것으로 판단할 수 있는 근거가 된다. 전체적인 득표율의 차이를 보기 위하여 <그림 5>에서는 각 시·도별로 실제 득표율과 출구조사 득표율의 차이를 비교해 보았다.

전라남도를 제외한 16개 시·도에서 더불어민주당은 출구조사 결과를 이용한 예측 득표율이 실제 득표율보다 낮게 나타났고, 전라남도과 전라북도를 제외한 15개 시·도에서 미래통합당은 출구조사 결과를 이용한 예측 득표율이 실제 득표율보다 높게 나타나고 있다. 평균적으로 더불어민주당은 1.44%p 과소 예측되었고, 미래통합당은 1.32%p 과대 예측되었다.

김영원·최윤정(2011)은 보수성향이 강한 지역에서는 대체적으로 보수지향 후보자의 득표율이 과대 추정되는 현상이 발생하고, 진보성향이 강한 지역에서는 반대로 진보지향 후보자의 득표율이 과대 추정된다는 사실을 과거 국회의원 선거결과를 이용하여 통계적으로 확인했다. 또한 이준웅(2001, 2004)에 의하면, 유권자는 여론 환경이 자신이 지지하는 정당에 대해 불리하다고 인식하는 경우 정치적 의견표명을 삼가는 경향이 있는데, 이러한 경향성은 특히 보수적 유권자의 경우 더욱 강조되어 나타난다고 하였다. 하지만 이번 총선에서는 이런 경향이 나타나지 않았다는 것을 알 수 있다.



<그림 5> 21대 총선 주요 정당의 실제 득표율-출구조사 득표율(단위: %)

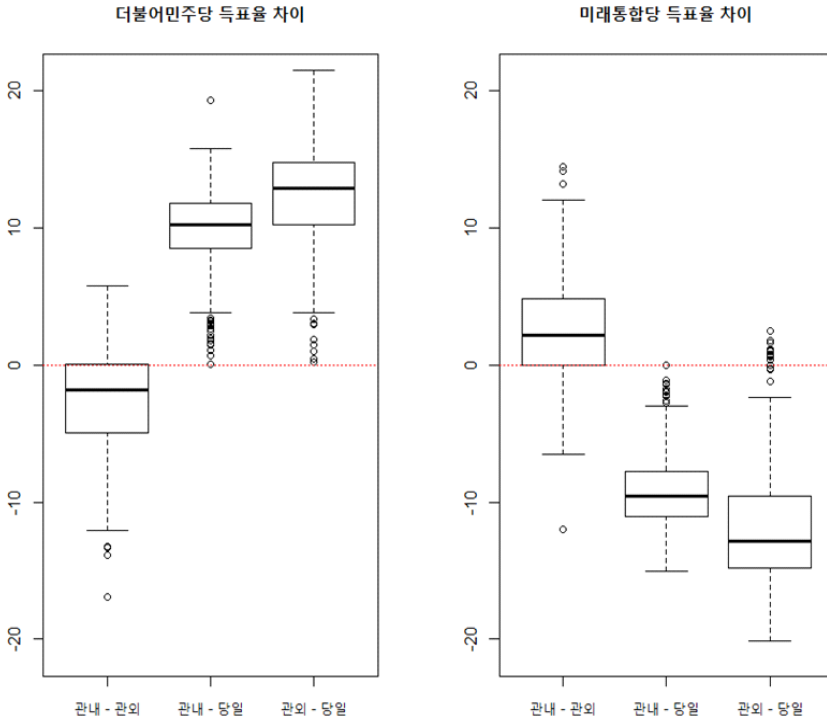


<표 7>을 보면 과거 국회의원 선거결과 예측을 위하여 사용한 무응답 그리고 5시 이후 투표자를 반영하기 위한 보정방법이 당일 득표율의 예측을 위해 적절하게 작동하였음을 알 수 있다. 그러나 당일 출구조사를 기반으로 한 사전투표 결과의 예측에는 심각한 오류가 발생하였으며, 이는 사전투표자의 성과 연령 기준 분포를 이용한 조정이 타당하지 않음을 나타낸다. 특별히 연령 보정의 효과를 살펴보기 위하여 <표 8>에서는 21대 총선에서의 사전투표자의 비중 및 연령 구성비를 살펴보았다. 유권자 분포와 사전투표자의 구성비는 2020 총선 출구조사 평가회의 자료에서 획득하였으며, 전체 투표자의 구성비는 중앙선거관리위원회의 연령대별 투표율을 2020년 인구추계자료에 적용하여 계산하였다. 연령 그룹별 투표율의 차이는 일부 존재하나, 사전투표와 당일투표의 연령 그룹별 차이는 거의 발생하고 있지 않음을 확인할 수 있다. 즉, 연령을 고려하더라도 사전투표자와 당일투표자 사이에 투표 성향에 체계적인 차이가 있고, 따라서 사전투표자에 대한 별도의 조사가 없이 출구조사 결과만을 이용한 예측에는 한계가 있음을 알 수 있다.

이러한 결과는 과거 선거와는 다른 투표행위가 제21대 국회의원 선거에서 이루어졌음을 의미하며, 그 원인은 사전투표자의 투표행위가 성과 연령을 고려한 후에도 체계적인 차이일 것으로 예측되며, 이를 살펴보기 위하여 과거 국회의원 선거와 그 차이가 명확하게 나타난 사전투표의 영향력을 살펴보았다.

<표 8> 21대 투표자 비중 및 연령 구성비(%)

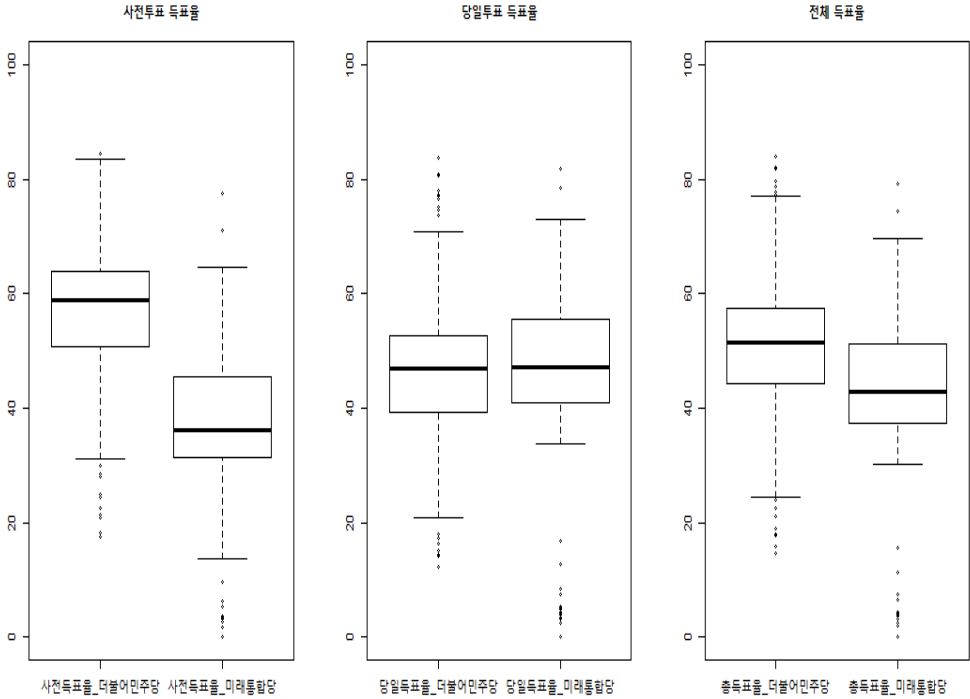
	유권자	전체 투표자	사전투표자
투표자 비중	-	100.0	40.3
~29세	18.1	18.4	16.9
30~39세	15.9	13.7	12.7
40~49세	19.0	17.6	17.7
50~59세	19.7	20.4	21.9
60~69세	14.6	17.3	18.3
70세~	12.7	12.6	12.4



<그림 6> 21대 총선 주요 정당의 투표유형별 득표율 차이(단위: %p)

현행 공직선거법에서는 사전투표 날에 출구조사가 불가능하다. 따라서 이번 총선 처럼 전체 투표자 중 40%가 넘는 사전투표자의 투표성향이 당일 투표자의 투표 성향과 다르다면, 출구조사만을 통한 선거결과 예측은 가능하지 않다. 사전투표자와 당일투표자의 투표성향에 차이가 있는지 확인하기 위해 주요 두 정당의 투표유형별 득표율을 비교했다. 우선 사전투표는 해당 구·시·군위원회 관할구역 내에 주소를 두고 있는 선거인을 대상으로 하는 관내사전투표와 관할구역 밖에 주소를 두고 있는 선거인을 대상으로 하는 관외사전투표로 나뉜다. 관내사전투표와 관외사전투표 그리고 당일투표와의 관계를 알아보기 위해 득표율 차이를 <그림 6>에 나타내었다. 두 정당 모두 관내사전투표 평균 득표율과 관외사전투표 평균 득표율 차이는 2~3%p 정도로 크게 차이 나지 않았다. 하지만 관내사전투표와 당일투표, 관외사전투표와 당일투표를 비교했을 때 평균 득표율 차이는 8~12%p 정도로 크게

나타난다. 이를 통해 사전투표와 당일투표는 구분하여 다룰 필요가 있지만, 관내사전투표와 관외사전투표는 특별히 구분할 필요가 없다고 판단하였다.



<그림 7> 사전투표 득표율, 당일투표 득표율, 전체 득표율의 비교(단위: %)

이후 사전투표 득표율, 당일투표 득표율, 그리고 전체 득표율의 관계를 <그림 7>을 통해 확인하였다. 당일투표에서 더불어민주당의 평균 득표율은 46.06%, 미래통합당의 평균 득표율은 44.78%로 거의 차이가 없었다. 하지만 사전투표에서 더불어민주당은 56.35%, 미래통합당은 35.52%의 평균 득표율을 보이며 20%p 이상의 차이를 보였고, 이 차이로 인해 전체 평균 득표율에서 더불어민주당 50.18%, 미래통합당 41.11%를 기록했다. 이를 통해, 사전투표 집단과 당일투표 집단의 투표 향에 명확한 차이가 있음을 확인하였다.

이러한 현상은 사전투표를 연구한 최근의 문헌에서도 언급되었다. 가상준(2018)은 사전투표하는 유권자의 특징이 공식 선거일에 투표하는 유권자와 점차 달라짐에

따라 공식 선거일에 투표하는 유권자만을 조사하여 선거결과를 예측하는 것은 불가능할 것이라고 하였다. 김찬송 외(2016)는 20대 총선에서 사전투표에 참여한 유권자들이 공식투표일에 참여한 유권자들과 선거 관심도, 정치적 태도, 연령, 직업 등에 차이가 있음을 확인했다.

제21대 국회의원 선거결과와 출구조사 결과의 차이가 발생한 원인은 과거와 매우 다르게 나타나고 있다. <그림 7>에서 본 바와 같이 출구조사 결과는 당일투표한 유권자들의 실제 결과를 매우 정확하게 반영하고 있으며, 따라서 무응답 효과, 보수 지지자 효과 그리고 열세자 효과 등은 나타나지 않는 것으로 판단할 수 있다. 제21대 국회의원 선거에서 발생한 예측 오차에 대한 보다 타당한 원인으로는 전체 투표자 중 사전투표자의 비중이 상대적으로 과거 선거보다 높은 상황에서 투표 당일에 조사된 투표자만의 행위를 기반으로 전체 선거결과를 예측하는 출구조사의 한계를 지적할 수 있다. 즉, 제21대 국회의원 선거결과에의 예측을 위한 출구조사에서는 사전투표자를 포함한 실제 투표자 전체에 대한 대표성을 갖추지 못한 추출틀 오차(frame error)로 인하여 예측에 체계적인 오류가 발생한 것으로 판단할 수 있다.

본 연구결과와 더불어 기존 연구결과를 통해 볼 때 사전투표자와 당일투표자의 정치적 성향에 체계적인 차이가 있는 것을 확인할 수 있었으며, 보다 정확한 예측을 위해서는 조사방법론 측면과 통계적 추정에 대한 추가적인 연구가 필요할 것으로 판단된다.

### Ⅲ. 결론 및 추후 연구

본 연구에서는 출구조사를 통한 21대 국회의원 선거 의석 수 예측의 실패 원인을 분석하였다. 지역별 당선자 수와 투표율을 살펴보고 대칭형 대응분석을 통해 지역, 사전투표 여부와 정당 지지율의 관계를 살펴보았다. 그리고 최근 3번의 국회의원 선거에서 최다 득표 정당, 경합지역, 예측 실패 시·군·구를 지도에 시각화하여 비교하였다. 21대 총선은 제1당과 제2당의 의석 수가 가장 많이 차이 났지만, 상대적으로 경합 지역이 많았던 것을 확인할 수 있었다.

의석 수 예측실패는 기본적으로 출구조사 자료 기반의 추정량이 갖는 체계적 편

향 때문이다. 예측에 실패한 14곳의 선거구 중 11곳에서 미래통합당을 예측했지만 더불어민주당이 당선되었다. 그리고 출구조사 결과가 대부분의 시·도에서 더불어민주당 후보자의 지지율은 과소 예측하고 미래통합당 후보의 지지율은 과대 예측한 것을 확인하였다. 그리고 예측에 실패한 선거구 중 1, 2위 득표율 차이가 1%p 이하인 초경합 지역은 한 곳으로, 단순히 경합 때문에 예측에 실패한 것이 아니라 출구조사 결과에 근거한 추정량의 체계적인 편향이 예측실패의 원인으로 작용한 것을 알 수 있다.

출구조사 결과가 갖는 체계적인 편향의 원인을 보다 상세히 살펴보기 위하여 사전투표와 당일투표 결과를 비교하였다. 관내사전투표와 관외사전투표는 비슷한 경향을 보이지만, 관내사전투표와 당일투표, 관외사전투표와 당일투표는 평균 8~12%p의 득표율 차이를 보이며 그 성격이 다른 것을 확인하였다. 그리고 사전투표 집단과 당일투표 집단의 정당 득표율 비교를 통해 두 집단의 투표성향에 차이가 있음을 확인했다.

즉, 당일투표자를 대상으로 이루어지는 출구조사 결과를 이용하여 선거결과를 예측함에 있어 제21대 국회의원 선거에서는 그 편향의 원인이 전통적으로 논의되었던 열세자 효과나 진보/보수 효과에 있는 것이 아니라, 사전투표 결과 예측의 실패에 그 원인이 있는 것으로 파악된다. 이는 사전투표자와 당일투표자 사이에 체계적인 차이가 있음을 의미하고 따라서 당일투표자만을 대상으로 한 선거결과 예측에는 한계가 있음을 의미한다.

앞으로의 선거 예측에서 이번 총선처럼 사전투표율이 높게 나타나고, 기존 출구조사의 문제점들이 사전투표자와 당일투표자의 체계적인 차이에 포함되어 나타난다면 출구조사 결과만을 이용한 예측은 힘들어질 것이다. 이를 보완할 수 있는 예측 방안에 대한 연구의 방향은 먼저 사전투표자를 대상으로 한 출구조사의 확대를 고려할 수 있다. 그러나 사전투표자 대상 출구조사는 비용과 시간 같은 현실적인 제약뿐 아니라 법적 근거를 요구하기 때문에 단기간에 이를 적용하는 것은 매우 어렵다. 또 다른 방안으로 사전투표자의 의견이 반영된 선거 이전에 이루어지는 다양한 여론조사 결과를 활용하는 방안을 고려할 수 있을 것이다. 이를 위해서는 21대 국회의원 선거 전에 이루어진 여론조사 결과에 대한 검토 및 이를 활용한 선거결과 예측을 위한 통계적 방법론에 대한 연구가 필요할 것으로 판단된다.

## 참고문헌

- 가상준. 2018. “사전투표 유권자의 특징 변화.” 《한국정당학회보》 17(4): 99-120.
- 김영원·최윤정. 2011. “출구조사의 체계적인 예측 편향에 대한 분석: 2010년 지방선거 출구조사를 중심으로.” 《조사연구》 12(3): 25-48.
- 김찬승·유재승·이현우. 2016. “사전투표자 세부분석.” 《21세기정치학회보》 26(4): 1-23.
- 박민규·윤상영. 2019. “국회의원 선거에서의 체계적 예측 편향을 고려한 주요 정당 의석 수 예측.” 《조사연구》 20(1): 29-43.
- 이준웅. 2001. “여론환경에 대한 인식과 정치적 의견 표명: 제16대 총선 예측조사 실패에 대한 정치 커뮤니케이션적 논의.” 《한국방송학보》 15(1): 199-236.
- 이준웅. 2004. “제17대 총선 예측 조사의 문제: 조사거절자와 응답불성실자 편향을 중심으로.” 《언론정보연구》 41(1): 110-135.
- 허명희. 2008. “국회의원 선거에서의 주요정당 의석 수 예측.” 《조사연구》 9(1): 87-100.
- 허명희. 2014. 《응용데이터분석》(1판) (pp. 203-220). 자유아카데미.
- “대한민국 최신 행정구역(SHP) 다운로드” (2020년 8월 19일) GIS DEVELOPER. 2020년 6월 15일 수정.  
<http://www.gisdeveloper.co.kr/?p=2332>.

<접수 2020.12.17; 수정 2021.01.15; 게재확정 2021.01.19>

## **An Analysis of Statistical Methods in Predicting the Number of Seats in the 21st General Election**

Seung-Ho Cheon

(Korea University)

Yo-Han Lim

(Korea University)

Min-Gue Park

(Korea University)

In the 21st general election held in April 2020, predictions of the number of seats based on exit polls is not successful. In this study, we investigated the problems in predicting the number of party seats using exit polls. To this end, we analyzed the results of the 21st general election in various perspectives and examined the differences between the 21st general election and the previous general elections. In the 21st general election, early voting rate was highest since it has been introduced, and a significant bias is introduced because only exit polls results were used for prediction. That is, there was a frame error that fails to cover the entire voter.

Key words: general election, exit poll, systematic bias, Bayes inference, early voting