

## 방법론 설명

본 사이트에 게재된 대통령 및 정당 지지율 추정값은 Simon Jackman(2005)의 논문 “Pooling the polls over an election campaign” (*Australian Journal of Political Science*) 에 기술된 방법론을 적용하여, 각 조사업체가 가진 경향성을 감안하여 추정한 결과입니다.

위 논문의 저자는 개별 여론조사에서 나오는 지지율 결과만으로는 선거 기간 여론의 세밀한 변화를 추적하기에 한계가 있다고 보고, 여러 조사기관의 조사 결과를 합치는 방법을 제시했습니다. 이때 각 조사기관의 특성이나 오차를 고려합니다. 어떤 조사기관은 특정 정당의 지지율을 전반적으로 높게 평가할 수도 있고, 어떤 조사는 오차가 조금 더 클 수도 있습니다.

풀에이에서 사용한 모델은 다음과 같습니다. 개별 여론조사 관측값( $y_i$ )은 여론 추세의 움직임( $\alpha_{t_i}$ )에 조사기관 효과( $\delta_{j_i}$ )와 조사오차에서 발생하는 오류( $\sigma_i$ )가 결합하여 결정된다고 가정합니다.

$$y_i \sim Normal(\mu_i, \sigma_i^2)$$

$$\mu_i = \alpha_{t_i} + \delta_{j_i}$$

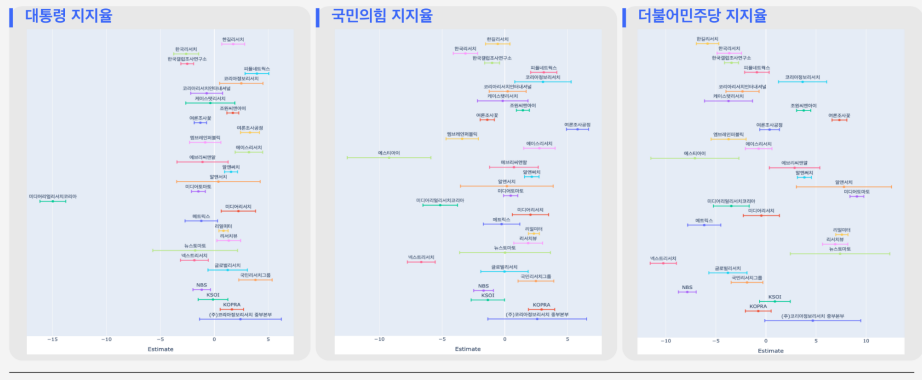
여기서 말하는 개별 여론조사 관측값이란 특정 시점에서 실시된 여론조사의 결과값을 의미합니다. 예를 들어 “○월 ○주 대통령 국정운영 지지율은 ○○%”와 같은 구체적인 숫자가 여기에 해당합니다. ( $i$ : 개별 여론조사 인덱스)

여론 추세의 움직임이란 시간이 지남에 따라 변화하는 여론의 경향을 말합니다. 대통령 지지율이 올라가고 있다거나, 특정 정당의 지지율이 떨어지고 있다는 식의 경향성을 의미합니다.

조사기관 효과, 즉 하우스 이펙트는 각 여론조사 기관이 가질 수 있는 특정한 경향성을 의미합니다. 하우스 이펙트는 아래와 같이 대통령 지지율, 정당별 지지율에서 각각 다른 분포를 보이기도 합니다.

## House Effect Analysis

- 플러스(+)는 평균 지지율보다 높게 추정하는 경향성 정도를, 마이너스(-)는 평균 지지율보다 낮게 추정하는 경향성 정도 ( $\mu_i = \alpha_i + \delta_j$ )
- 등록 취소 기관을 제외하면, 대통령, 국민의힘 지지율에 비해 민주당 지지율이 조사기관들의 과대 / 과소 추정 폭이 크게 나타남



마지막으로 조사오차에서 발생하는 오류는 여론조사를 할 때 발생할 수 있는 일반적인 오류나 불확실성을 의미합니다. 결국 우리가 접하게 되는 개별 여론조사 결과(관측값)는 위의 요소들이 결합되어 복합적으로 결정된다고 보는 것입니다.

다음으로 폴에이는 위의 모델에 '베이지언 추론'이라는 통계적 방법을 적용, 각 시점의 여론 추세의 움직임과 조사기관 효과를 추정하고자 했습니다. 구체적으로 마코프체인 몬테카를로 (Markov Chain Monte Carlo (MCMC)) 중 Gibbs sampling이라는 알고리즘을 사용했습니다.

모수의 사전분포로부터 사후분포를 업데이트해나가는 연쇄적인 추론 과정이라 볼 수 있습니다. 하나하나 차례대로 여러 변수의 값을 추정하면서 전체 그림을 만들어가는 셈입니다.

**1-1)** 지지율 추정값의 사전분포는 아래와 같이 정의했습니다.

$$\alpha_1 = (0.5, 0.5)$$

$$\alpha_t \sim \text{Normal}(\alpha_{t-1}, \omega_t^2), t = 2, \dots, T$$

$\omega^2$ : 일별 가변성의 크기를 나타내는  $\alpha_t$ 의 분산

**1-2)** 지지율 추정값의 사후분포를 구하는 과정은 다음과 같습니다.

1.  $\alpha_1$ 을 선택
2.  $\alpha_t(k+1) \sim p(\alpha_t(k) | \alpha_1(k+1), \dots, \alpha_{t-1}(k+1), \alpha_{t+1}(k), \dots, \alpha_T(k), Y)$
3.  $\alpha_t(k+1)$ 을 표집에 추가한다.
4. 위의 과정을 2,000번 반복하고, 앞의 1,000번의 반복은 제거한다.

**1-3)** 지지율 추정값의 신뢰구간은 위 사후분포에서 추출됩니다.

**2-1)** 조사기관 효과의 사전분포는 아래와 같이 정의했습니다.

$$\delta_j \sim \text{Normal}(0, d^2)$$

$j$  = 각 조사기관 인덱스,  $d$  = 임의의 큰 상수

**2-2)** 지지율 추정값의 사후분포는 **1-2)**와 같은 방법으로 반복합니다.

**2-3)** 조사업체 효과 추정값의 신뢰구간은 위 사후분포에서 추출됩니다.