

2022年参議院選挙におけるインターネット調査と予測手法の検討

Internet Survey and Prediction Method for the 2022 House of Councillors Election

福田 昌史
Masafumi Fukuda



1. 序
2. 調査の概要
3. MRPの適用
4. 結果
5. 議論

〈要旨〉

読売新聞社とIT大手ヤフー（現LINEヤフー）は2022年参議院選挙において、インターネットを通じた情勢調査を試験的に実施した。16選挙区で実施した調査データを基にMRP (Multilevel Regression and Poststratification) を適用し、候補者の得票率を予測したところ、予測値と得票率の平均絶対誤差に一定の改善がみられた。

The Yomiuri Shimbun and Yahoo corp. (Line Yahoo) conducted an experimental internet survey for the 2022 House of Councillors election. Multilevel Regression and Poststratification (MRP) was applied to the survey data collected in 16 electoral districts and the results showed a certain improvement in the mean absolute error between the predictions and the share of the vote of the candidates.

1. 序

国政選挙における報道機関の情勢調査は、現在、調査手法の変化の中にある。従来は調査員による電話調査が主流だったが、最近の選挙では自動音声による調査やネット調査が使われるようになった。2019年の参議院選挙では、毎日新聞社が自動音声による電話調査を導入し、2021年の衆議院選挙では、共同通信社が自動音声の調査を、朝日新聞社がネット調査を取り入れた。この選挙で読売新聞社は、日本経済新聞社と自動音声の調査を共同で行った。さらに毎日新聞社は、2022年の参議院選挙でネット調査を行った。ネット調査は、ネット調査会社が調査のために集めた会員（モニター）に対する調査（朝日）や、携帯電話会社のポイント会員を対象としたもの（毎日）など、それぞれ形態は異なっている。

選挙の前に行う情勢調査は、通常、回答データを加工せずに集計すると選挙結果とは一致せず、一定の傾向を持ってズレが生じている。その要因として、回答者の中には投票に行かない人が含まれていることや、投票に行くとしても調査時点では投票先を決めていない人などが含まれていることなどが考えられる。

このため、報道機関は過去における選挙の調査結果と実際の選挙結果の関係を分析し、予測モデルを作成、このモデルを次の選挙の調査データに適用することで、候補者の当落可能性や政党の獲得議席を推定してきた（例えば松田，2018）。それゆえ、同じ方式で実施した過去の選挙の調査データを蓄積していくことが非常に重要となる。

本稿は、2022年の参議院選挙（7月10日投開票）で、読売新聞社がIT大手のヤフー（現LINEヤフー）と共同で試験的に実施したネット調査（読売新聞，2022）と、そのデータを使った予測手法の検討について論じる。

2. 調査の概要

ネット調査は、参院選の投開票日約1週間前の7月1～2日の2日間、「Yahoo! JAPAN」のIDと、キャッシュレス決済「PayPay（ペイペイ）」のIDを連携しているユーザー（全国で約3700万人）を対象に実施した。選挙区の調査は、東京、大阪、沖縄など16選挙区（表1）に住むユーザーから選んで調査協力を呼びかけるメールを送り、ウェブ上に設けたアンケートに回答してもらった。全国が対象となる比例選は、ヤ

フーのアンケートサイトに案内を掲示して協力を募った。いずれも、回答してくれたユーザーには、ペイペイのポイントを進呈した。

調査対象の選挙区は、青森など改選定数1の1人区から東京（改選定数6）などの複数区まで幅広く含まれるように選び、回収目標は各選挙区1000人とした。

調査では、「選挙区選の投票先」、「比例選の投票先」を聞いたほか、「参院選への関心」「投票に行くか」「どの政党を支持しているか」「岸田内閣を支持するか」など約10問を尋ねた。

選挙区選の投票先を聞いた質問で、得られた集計値と実際の選挙結果を比べて当落の的中度をみると、16選挙区の当選者35人のうち32人が的中。このうち11選挙区では当選者の順位も一致した（表1）。

表1. 調査を実施した16選挙区

	北海道
	青森
	福島
	新潟
	山梨
当選者が的中（順位も調査結果と一致）	静岡
	三重
	兵庫
	鳥取・島根
	愛媛
	福岡
当選者は的中（順位が調査結果と異なる）	東京
	大阪
当選者が選挙結果と食い違う	埼玉
	京都
	沖縄

調査結果はおおむね選挙情勢を把握できる精度だったが、特定の政党に関する支持の強弱の「くせ」が現れた。京都選挙区（改選定数2）では、日本維新の会の候補が調査では1位で当選の枠内だったが、実際には3位で落選した。京都も含めた維新の候補7人は、実際の選挙結果より平均して6ポイント高く得票すると調査結果だった。埼玉選挙区でも維新候補の予測は「当選」だったが、実際は落選した。このほか、公明党の候補の支持が低めとなる傾向もあった。

回収サンプルの属性構成をみると、調査データは、女性より男性の回答が多いなどの偏りもみられた。

3. MRPの適用

本稿では、近年欧米で選挙予測に用いられることが増えているMRPを今回の調査データに適用し、選挙区候補の得票率の予測が改善されるかどうかを分析、検証する。MRPは、Multilevel Regression and Post-stratificationの略で、文字通り、マルチレベル回帰分析と事後層化を組み合わせた手法である。

MRPを選挙予測に使う場合、大まかにいうと、

(1) 性別や年代、地域などの属性や支持政党などの変数から特定の候補者に投票する確率を計算する統計モデルを作る

(2) (1)で作ったモデルを各選挙区の投票者全員に当てはめるとどうなるかを推定するという2段階を経る。

例えばWang, et al. (2015)は、ゲーム機X-Boxのユーザーを対象に2012年米国大統領選挙に関する連続調査を行い、サンプルの65%を18~29歳が占め、男性が93%を占めるなど、大きく偏ったサンプルをMRPによって補正し、実際の選挙結果に近い予測ができることを示した。ここでは投票者を17万以上のカテゴリー（性別=2、人種=4、年代=4、教育=4、州=51、政党支持=3、イデオロギー=3、2008年大統領選の投票先=3の組み合わせ）に分類して予測モデルを作った。

Kiewiet de Jonge et al. (2018)は、2016年の米国大統領選挙における州レベルの得票率予測にMRPを使った。投票者を州、性別、人種で408カテゴリーに分類してモデルを作り、51州（ワシントンDC含む）のうち50を正確に予測した。また、イギリスの調査会社YouGovは、2016年のEU離脱を巡る国民投票や、2017年、19年の総選挙でMRPを使って選挙を予測した（例えばYouGov, 2017）。

国内の情勢調査では、多くの質問を聞くことが一般的ではない。今回の検討では、1段階目のモデル推定に、性別（2カテゴリー）、年代（6カテゴリー）、政党支持（11カテゴリー）のみ（計132カテゴリー）を使った。45選挙区であることを踏まえると、サンプル全体での分類の数は $132 \times 45 = 5940$ となる。

2段階目の事後層化には、ターゲットとなる投票者データとして、この選挙の当日出口調査のデ

ータを加工して用いた。具体的には、集計が選挙区選の開票結果と比例選の開票結果に合うようにウェイトを付け直した（出口調査の投票先の回答はMRPの段階では使わない）。実際の選挙では、情勢調査データから選挙結果を予測する時点ではその選挙の出口調査は実施しておらず、今回のような予測を行うことは不可能である。そのため、この分析は、仮に理想に近いベンチマークデータが得られた場合に、MRPによってどの程度予測値が改善できるかを検証するものとなる。

分析には、実際に得られたデータから性別、年代、支持政党、選挙区の投票先で無回答だったケースを除外した、約7900件を使用した。また、予測の対象は、16選挙区の130人の候補のうち、主要政党の候補者のみ（自民、立民、維新、公明、国民、共産、れいわ、社民、政党推薦を受けた一部の無所属）、計67人の得票率を予測した。

4. 結果

各候補の予測値である推定得票率は、その選挙区の候補者の推定得票率の合計が100%となるようリサイズした。予測の評価については、候補者*i*の実際の得票率を p_i 、推定得票率を \hat{p}_i として候補者の平均絶対誤差（Mean Absolute Error）

$$MAE = \sum_{i=1}^n |\hat{p}_i - p_i| / n$$

を、MRP適用前と適用後で計算して比較した。MAEは、67候補全員では適用前の2.9から2.1に縮小された。

図1、図2は、それぞれ67人の候補の適用前と適用後の推定得票率（水平軸）、実際の得票率（垂直軸）を布置したものである。適用前と比べて適用後は、点のばらつきが斜め45度の直線にある程度近づいていることが分かる。

MAEの値を候補者の政党別に表2にまとめた。日本維新の会の候補は適用前の6.8ポイントから1.9ポイントに大きく改善がみられ、公明党候補についても5.8から3.9に低下した。その一方で、自民党候補は2.8から3.1に、立憲民主党候補は1.8から2.2に増え、改善がみられなかった。

5. 議論

今回の分析では、ネット調査のデータにMRPを適用し、一定の偏りをもった非確率サンプルで選挙予測の精度がどの程度改善されるかを検討し

た。その結果、平均絶対誤差について一定の改善がみられ、偏りが大きかった日本維新の会の候補については予測を大幅に改善できた。ただ、当落的中数が増えたわけではなく、さらなる改善が必要であるといえる。MRPを改善するには、第一段階の統計モデルの説明変数を増やしたり、第二段階の事後層化の粒度を細かくするなどが考えられる。

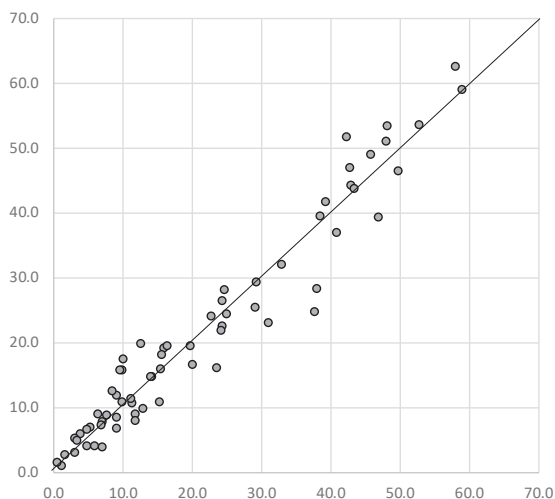


図1. MRP適用前の推定得票率（水平軸）と選挙結果（垂直軸）

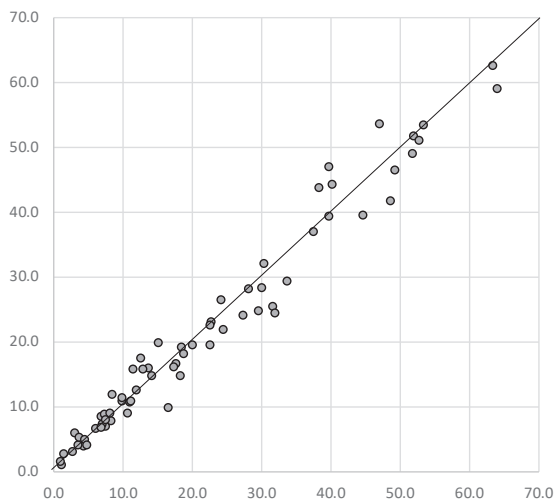


図2. MRP適用後の推定得票率（水平軸）と選挙結果（垂直軸）

さらに、この分析では事後層化のベンチマークとして、その選挙の出口調査データを活用した。先に触れたとおり、実際の情勢調査は、出口調査

が行われる投票日より前に実施するため、このようなことはできない。実際の場面で良好なベンチマークとなる投票者の属性構成データをどう作るかが予測の成否に関わる鍵となる。

表2. MAEの比較（パーセントポイント）

候補者の政党	適用前	適用後
自民党 (18)	2.8	3.1
立憲民主党 (16)	1.8	2.2
日本維新の会 (7)	6.8	1.9
公明党 (5)	5.8	3.9
共産党 (9)	1.6	0.5
国民民主党 (5)	2.2	1.1
れいわ新選組 (5)	2.0	1.6
社民党 (2)	0.6	0.3

※カッコ内は候補者数（推薦候補を一部含む）

また、今回の分析ではMRPのみを適用したため、単純な事後層化など他の手法より優れているかどうかは評価できていない。これらの検証も今後の課題といえる。

（読売新聞東京本社世論調査部）

参考文献

- Kiewiet de Jonge, Chad P., Gary Langer, and Sofi Sinozich (2018). Predicting State Presidential Election Results with National Tracking Polls and multilevel regression with poststratification (MRP). *Public Opinion Quarterly*, 82-3, 419-446.
- 松田 映二 (2018). 電話調査導入や選挙制度改変による選挙予測モデルの変遷, *行動計量学*, 2018, 45-2, 95-106.
- 読売新聞 (2022). 当選者の的中率91%—読売・ヤフー参院選ネット情勢調査, *読売新聞* 2022年12月7日朝刊.
- YouGov (2017). How the YouGov model for the 2017 General Election works, <https://yougov.co.uk/news/2017/05/31/how-yougov-model-2017-general-election-works/>, (2023年10月1日閲覧).
- Wang, W, David, R, Sharad, G and Gelman, A (2015). Forecasting elections with non-representative polls, *International Journal of Forecasting*, 31-3, 980-991.