

インターネット調査による選挙予測の可能性

Possibility of Election Prediction through an Internet Survey

江口 達也
Tatsuya Eguchi

1. はじめに
2. 今回実施したネット調査の設計
3. 「比例区投票先」のトレンドと過去データとの比較
4. ネット調査で比例区の予測は可能か
5. 傾向スコア補正による選挙予測の可能性
6. おわりに

<要旨>

近年、現行のRDD方式による電話調査では有権者の投票動向を十分に捉えられていない疑いが強い。電話調査を代替する手法の選択肢としてインターネット調査が考えられる。朝日新聞社が2012年と2014年の衆院選で実施したネット調査の結果をみると、各政党の趨勢を捉えることができおり、選挙予測に役立つ知見を与えてくれる可能性がある。さらに、2006年の長野県知事選で実施した郵送・ネット並行調査のデータを用いて「傾向スコア補正」を試行し、2つのネット調査うちの1つが郵送調査の結果に補正できることを確認した。

In recent years, telephone surveys that use the random digit dialing (RDD) method have been strongly suspected of leading to incorrect results about voting behavior. We consider that Internet surveys serve as an alternative to telephone surveys. The results of Internet surveys conducted by the Asahi Shimbun in the House of Representatives elections in 2012 and 2014 showed successful estimation and capability of this useful survey mode for election prediction. In addition, when we use the parallel data collected through both mail and Internet surveys in Nagano gubernatorial election in 2006 and apply "propensity score adjustment" in the prediction, the answer is equal to the election result for a case of two experiments.

1. はじめに

2014年12月に行われた衆議院選挙において、各新聞社および通信社はRDD方式の電話調査による選挙情勢調査を実施した。図表1は、各社の比例区予測議席である。議席推計表が紙面に掲載されている場合はその数字を、掲載されていない場合は記事中の表現を引用した。

図表1. 2014年衆院選 各社の比例区予測

政党名	結果	朝日新聞(中盤)	共同通信(序盤)	毎日新聞(中盤)	読売新聞(序盤)	日経新聞(終盤)
自民	68	70~ 76 ~82	77~ 82 ~87	74 ~ 79	70議席台となる勢い	有力圏内で70議席を超えるが、77議席を上回るのは難しそう
民主	35	34~ 39 ~43	31~ 35 ~40	35 ~ 42	30議席は上回るものの伸びは限定的	固めたのは30議席台半ば
維新	30	20~ 23 ~26	21~ 24 ~26	19 ~ 24	20議席台の情勢	20議席台にとどまりそう
公明	26	21~ 24 ~26	21~ 24 ~26	24 ~ 27	比例選で伸長し、全体として公示前の31議席を上回りそう	比例代表では議席を上積みし、全体で31議席から伸びそう
次世代	0	0~ 0 ~2	0~ 0 ~1	0	小選挙区選と比例選を合わせても、1けた台に	議席獲得の可能性が出てきた
共産	20	15~ 17 ~19	12~ 14 ~17	15 ~ 19	公示前8議席から倍増をうかがう	15議席を固め、さらに上積みを目指す
生活	0	0~ 0 ~1	0	0		
社民	1	0~ 1 ~2	1~ 1 ~1	0 ~ 1		
改革	0	0	0	0		
諸派	0	0	0	0		

公示日から実施された「序盤調査」と、投票日に近い時期に実施された「中盤(終盤)調査」という調査時期の違いはあるが、いずれも自民党を多めに、維新の党を少なめに予測してはざしている。また、共産党もやや少なめの予測となっている。

とくに、維新の議席予測がはずれるのは、2012年衆院選、2013年参院選に続き3回連続である。今回は小選挙区の議席予測でも、朝日新聞社を含め各社が維新の議席数を少なめに予測してはざしている。なぜ各社は維新を過小評価してしまったのか。予測式の精度という問題はあるにせよ、各社が収集したデータは「維新には前回ほどの勢いはない」としか読み取れないものであったことは共通しているのではないかと。維新投票層の動向を十分に捉えられないままの電話調査(RDD方式)を、補完もしくは代替する手段を探る必要性を強く感じている。

補完あるいは代替する手段の候補としてインターネット調査(以降、ネット調査)が考えられる。朝日新聞社は、2012年衆院選からインターネット調査を並行実施しており、今回も引き続き行った。本稿ではその結果の一部を紹介するとともに、ネット調査を選挙予測に活用できる可能性があるのかどうか検討する。

2. 今回実施したネット調査の設計

今回実施したネット調査は、2012年衆院選時と同様に、主にパソコンを通じて回答する「PCモニター」のうち全国の20歳以上の人を対象にした「PC調査」と、携帯電話を通じて回答する「モバイルモニター」のうち全国の20～39歳の人を対象にした「モバイル調査」の2種類の調査を実施した。PC調査、モバイル調査ともに、図表2のように性別と年齢を割り付けて「スクリーニング調査」を実施し、PC調査は10000人、モバイル調査は7426人を回収した。この回答者を性別、年齢、居住地、内閣支持、政党支持の回答内容で

図表2. サンプル割り付け

【PC調査】			【モバイル調査】		
	男性	女性		男性	女性
20～24歳	3%	3%	20～24歳	11%	10%
25～29歳	3%	3%	25～29歳	12%	11%
30～34歳	4%	4%	30～34歳	13%	13%
35～39歳	4%	4%	35～39歳	16%	15%
40～44歳	5%	5%			
45～49歳	4%	4%			
50～54歳	4%	4%			
55～59歳	4%	4%			
60歳以上	18%	22%			

層別した上で系統抽出し、似通った2000人のグループをそれぞれ3つずつ作成した。調査はPC調査、モバイル調査ともに投票日前に3回、系統抽出した各グループに対して順に調査を依頼する方式で行った。日程および有効回答数については図表3の通り。

図表3. 調査日程と有効回答数

	日程	有効回答数	
		PC	モバイル
スクリーニング	2014年11月28日～12月1日	10000	7426
投開票前①	2014年12月6日～7日	1649	1453
投開票前②	2014年12月9日～10日	1612	1389
投開票前③	2014年12月12日～13日	1550	1375

3. 「比例区投票先」のトレンドと過去データとの比較

比例区投票先を聞いた結果を図表4にまとめた。この数字は投票先を回答しなかった人を除いて計算した比率である。なお、PC調査とモバイル調査で回答者集団の年齢構成が異なることに注意する。

数字のトレンド（動き）を確認すると、PC調査、モバイル調査ともに自民は投票日直前の第3回調査で支持率が減少している。また、維新はPC調査で18%→19%→21%、モバイル調査で14%→16%→17%と、ともに微増傾向を示している。朝日新聞社が実施した序盤と中盤の情勢調査（RDD方式）でも比例区投票先

図表4. 比例区投票先

	PC調査			モバイル調査		
	投票前①	投票前②	投票前③	投票前①	投票前②	投票前③
自民	37%	37%	34%	46%	46%	40%
民主	16%	16%	15%	12%	11%	12%
維新	18%	19%	21%	14%	16%	17%
公明	5%	5%	5%	7%	7%	8%
次世代	3%	4%	4%	5%	4%	5%
共産	14%	14%	14%	8%	8%	9%
生活	2%	2%	3%	2%	2%	2%
社民	3%	1%	2%	2%	1%	2%
改革	0%	1%	0%	1%	1%	1%
幸福	2%	1%	1%	3%	3%	4%

では自民は減少し、維新は微増している。RDD 調査とネット調査が同じトレンドであること、選挙結果が情勢調査に基づく予測よりも自民が少なめ、維新は多めとなったことを考慮すれば、ネット調査でも選挙期間中の各政党の趨勢を十分に捉えられていたといえるのではないかな。

では、今回衆院選での維新の「勢い」は前回衆院選に比べて衰えていたのだろうか。2012年衆院選で行ったネット調査の結果と比較することで確認してみる。図表5に2012年衆院選ネット調査で比例区投票先を聞いた結果をまとめた。投票日に最も近い前回第6回調査と今回第3回調査の維新支持率を比較すると、若年層が対象のモバイル調査では22%→17%とやや勢いに衰えがみられるものの、全年代が対象のPC調査では23%→21%とそれほど衰えていない。RDD調査とは違うネット調査の動向が、選挙予測に役立つ知見を与えてくれるのではないかな。

図表5. 2012年衆院選ネット調査の比例区投票先

	PC調査					
	投票前①	投票前②	投票前③	投票前④	投票前⑤	投票前⑥
民主	13%	13%	14%	12%	13%	13%
自民	29%	29%	27%	29%	29%	27%
未来	-	8%	8%	6%	7%	6%
公明	3%	4%	5%	4%	4%	5%
維新	27%	26%	24%	29%	23%	23%
共産	5%	5%	6%	5%	7%	7%
みんな	10%	12%	11%	11%	12%	15%
社民	2%	1%	2%	2%	2%	1%
大地	1%	1%	1%	1%	1%	1%
国民	0%	0%	1%	0%	0%	1%
日本	0%	0%	1%	1%	1%	1%
改革	2%	1%	1%	1%	2%	1%
生活第一	5%	-	-	-	-	-
減税	2%	-	-	-	-	-
みどり	1%	-	-	-	-	-

	モバイル調査					
	投票前①	投票前②	投票前③	投票前④	投票前⑤	投票前⑥
民主	13%	15%	12%	14%	16%	14%
自民	28%	29%	28%	29%	31%	29%
未来	-	4%	5%	6%	5%	6%
公明	5%	5%	9%	6%	6%	6%
維新	25%	26%	26%	23%	22%	22%
共産	4%	5%	4%	5%	5%	5%
みんな	7%	11%	9%	10%	9%	11%
社民	1%	1%	2%	2%	2%	2%
大地	1%	2%	1%	1%	1%	1%
国民	1%	1%	1%	1%	1%	1%
日本	0%	0%	1%	1%	0%	1%
改革	2%	1%	2%	1%	2%	3%
生活第一	7%	-	-	-	-	-
減税	3%	-	-	-	-	-
みどり	1%	-	-	-	-	-

4. ネット調査で比例区の予測は可能か

では、ネット調査を使って選挙予測をするという観点でデータをみるとどうか。今回のネット調査は選挙区ごとではなく全国調査の形式で行っているため、ここでは比例区の予測可能性を考えることにする。また、モバイル調査は20-39歳に限定した調査であるため、PC調査の結果を基に考察する。

今回の選挙結果（選挙得票率）とネット調査の結果および前回衆院選の選挙結果とネット調査結果を比較してみる（図表6）。ここに示したネット調査の結果は投票日に最も近い時期に実施したPC調査のものである。

今回と前回で共通しているネット調査の特徴は、①維新の支持率が結果よりも多め、②公明の支持率が結果よりもかなり少なめ、③民主の支持率もやや少なめ、④自民の支持率は結果とかなり近い、といった点が挙げられる。仮にこうした傾向が常にネット調査に表れるのであれば、それに合わせた予測式が作成できるため、ネット調査を使って選挙予測ができそうである。

試しに 2012 年のデータを用いて単純な予測を行ってみよう。まず 2012 年衆院選比例区の各政党の選挙得票率を、ネット調査の支持率で除して各政党の「係数」を求める。例えば維新の係数は、 $20.4\% \div 23\% = \text{約} 0.9$ となる。2012 年に存在しなかった新党の係数は

1.0 とした。そして、求めた係数を 2014 年衆院選ネット調査の支持率に掛けて「推計得票率」を算出した（図表 7）。

2014 年衆院選の選挙得票率とネット調査支持率の誤差は、公明で-8.7、維新で+5.3、民主で-3.3 などとなっている。一方、選挙得票率と推計得票率の誤差は、公明-1.9、維新+2.9、民主+0.1 などと縮小している。自民のように+0.9→+1.7 と逆に拡大した政党もあるものの、補正前の残差二乗和は 126.1 だったが、補正後は 23.8 まで縮小できており、全体として補正はうまく機能している。

しかし、実際の予測は衆院比例ブロックごとに行う必要があるし、議席推計のためには推計得票率を当てることよりも、各党の獲得議席の大小関係を当てることの方がより重要になる。また、この選挙 2 回分のデータだけでは共通した傾向が“たまたま出た”という可能性を排除できない。引き続きデータを蓄積し、このような倍率補正方式の妥当性を検討していきたい。

図表 6. 選挙結果とネット調査の比較

■ 2014年衆院選			■ 2012年衆院選		
	選挙得票率	ネット調査 (PC③)		選挙得票率	ネット調査 (PC⑥)
自民	33.1%	34%	自民	27.6%	27%
民主	18.3%	15%	民主	16.0%	13%
維新	15.7%	21%	維新	20.4%	23%
公明	13.7%	5%	公明	11.8%	5%
次世代	2.7%	4%	みんな	8.7%	15%
共産	11.4%	14%	未来	5.7%	6%
生活	1.9%	3%	共産	6.1%	7%
社民	2.5%	2%	社民	2.4%	1%
改革	0.0%	0%	大地	0.6%	1%
幸福	0.5%	1%	国民	0.1%	1%
支持なし	0.2%	—	改革	0.2%	1%
			幸福	0.4%	—

図表 7. ネット調査から 2014 年選挙結果を予測

	ネット調査 (PC③)[A]	係数 [B]	推計得票率 [C]	選挙得票率 [D]	ネット調査との誤差 [A-D]	推計得票率との誤差 [C-D]
自民	34%	× 1.0 =	34.8%	33.1%	+0.9	+1.7
民主	15%	× 1.2 =	18.5%	18.3%	-3.3	+0.1
維新	21%	× 0.9 =	18.6%	15.7%	+5.3	+2.9
公明	5%	× 2.4 =	11.8%	13.7%	-8.7	-1.9
次世代	4%	× 1.0 =	4.0%	2.7%	+1.3	+1.3
共産	14%	× 0.9 =	12.3%	11.4%	+2.6	+0.9
生活	3%	× 1.0 =	3.0%	1.9%	+1.1	+1.1
社民	2%	× 2.4 =	4.7%	2.5%	-0.5	+2.3
改革	0%	× 0.2 =	0.0%	0.0%	-0.0	-0.0
幸福	1%	× 1.0 =	1.0%	0.5%	+0.5	+0.5
支持なし	—	—	—	0.2%	—	—

5. 傾向スコア補正による選挙予測の可能性

海外ではネット調査を使った選挙予測は既に行われている。最近では、イギリス総選挙で各世論調査会社が行った「保守党と労働党の大接戦」という予測が大きくはずれたことが話題となったが、その半数以上はネット調査を使って行われていた（2015年5月15日付読売新聞朝刊）。また、2000年のアメリカ大統領選でブッシュとゴアの得票がほぼ互角だと唯一当てたハリス・インタラクティブ社の予測はネット調査に基づくもので注目された。このときハリス社がネット調査の補正に使った「傾向スコア」という手法も同時に注目されるようになった。傾向スコアは海外のネット調査会社ではよく使われているようだが、日本でもビデオリサーチ社が実用化しているなど、マーケティング調査の世界では既に使われている手法である。だが、選挙予測の世界で実用化されたという例はまだ見あたらない。

朝日新聞社は2006年9月に行われた長野県知事選挙について、郵送調査とネット調査の並行調査を実施している。古いデータではあるが、これを使って傾向スコアを用いたネット調査の補正を試行した。

図表8. 村井・田中両氏の支持率

	選挙結果	郵送調査	ネット調査 (A社)	ネット調査 (B社)
村井仁氏	53.4%	52%	34%	34%
田中康夫氏	46.6%	48%	66%	66%

2006年9月の長野県知事選挙には、新人の村井仁氏と現職の田中康夫氏の2人が立候補し、村井氏が得票率53.4%を得て田中氏（得票率46.6%）を下して当選した。郵送調査、ネット調査で得られた両氏の支持率を図表8に示した。このとき、ネット調査は2つの会社に委託して行った（ここではA社、B社とした）。調査対象者は長野県内に住む20歳以上の登録モニターである。

郵送調査は村井氏がややリードしていることを的確に捉えており選挙結果にかなり近いが、ネット調査はA社、B社ともに落選した田中氏が大きくリードしているという結果になった。この相違はネット調査が登録モニターを対象としていることによる偏りに起因すると考えられる。この偏りを、傾向スコアを用いて補正することを試みる。

傾向スコア補正は、まず郵送とネットで共通した質問項目（共変量）を使い、ネット調査回答者の「ネット調査の回答者である確率」（傾向スコア）をロジスティック回帰によって推定する。次に、その確率が低い人の重みを増し、逆に確率が高い人の重みは減らすという方法で行った（逆確率による重み付け（IPW））。今回使用した共変量は、①性別、②年齢、③政治満足度、④生活満足度、⑤政党支持、⑥ネット利用頻度、⑦郵送調査への抵抗度、の7項目である。補正結果を図表9に示したが、この傾向スコア補正で予測している

のは「郵送調査回答者（つまり無作為抽出標本）がネット調査に回答した場合の結果」であって「選挙結果」ではないことに注意したい。

図表9. 傾向スコアによる補正結果

	A社		B社		郵送調査
	補正前	補正後	補正前	補正後	
村井仁氏	34%	→ 52%	34%	→ 39%	52%
田中康夫氏	66%	→ 48%	66%	→ 61%	48%

補正の結果、A社のデータでは村井氏の支持率が田中氏を上回り、郵送調査の結果にうまく補正できている。しかしB社のデータでは、村井氏と田中氏の差はやや接近したものの、田中氏のリードは変わらなかった。

今回使用した共変量は、星野・前田(2006)の提案に基づき、以下の考え方で選んだ。

- ①個人内変動が少なく、かつネット調査と郵送調査で継続的に質問できる可能性が大きい項目
- ②郵送調査とネット調査間で差のある項目
- ③補正の目的項目（ここでは知事選の投票先）を共変量に回帰させたときの偏回帰係数が、2群とも同じ方向になるものを選ぶ
- ④上記の基準で選択された共変量のセットから、さらに二乗誤差の和を減少させるように共変量を減らす

長野知事選での並行調査は傾向スコアによる補正を目的として実施されたものではなく、使用できる共変量の項目に限りがあったが、一定の補正効果を確認できた。ただ、同じ共変量を使用したにも関わらず、A社とB社で補正効果に大きな差があった。星野(2009)が指摘するように、共変量項目はモニター群ごとに検討する必要がある。また、もしA社（またはB社）に委託して長野県の別の選挙でネット調査を実施した場合、今回使った共変量が有効に機能するかは検証されていないことも付け加えておきたい。

6. おわりに

登録モニターを対象としたネット調査には、無作為抽出調査とは異なる独特の偏りが発生する可能性がある。だが、データを見てみると選挙予測に全く役に立たないと切って捨てるようなものだとは思えない。我々はその偏りを補正する術を持っていないだけではないか。無作為抽出調査にも偏りは当然発生するが、その補正を試みることで選挙結果の予測を行っている。同様のアプローチがネット調査に対しても行われて良いはずである。

本稿では補正手段として倍率補正方式や傾向スコアによる補正を試行した。先に述べた

ように傾向スコアによる補正は選挙結果を予測しているのではなく、あくまで無作為抽出標本がネット調査に回答した結果の予測である。だが、無作為抽出標本の回答結果が予測できれば、これまで蓄積してきたデータや予測手法を使った選挙予測が可能になる。推定を重ねるため誤差の拡大という精度面での不安はあるが、研究する価値はあると考える。傾向スコアに限らず、引き続きネット調査の活用法を検討していきたい。

(朝日新聞社 世論調査部)

〈参考文献〉

- 江口達也(2013a). 「衆院選ネット調査について」『サーベイ・メソドロジー研究会報告』, 2013. 3. 25.
- (2013b). 「ネット調査による若年層捕捉の可能性」, 『政策と調査』, 5, 19-27.
- 星野崇宏・前田忠彦(2006). 「傾向スコアを用いた補正法の有意抽出による標本調査への応用と共変量の選択法の提案」, 『統計数理』, 54(1), 191-206.
- 星野崇宏(2009). 「調査観察データの統計科学 因果推論・選択バイアス・データ融合」, 岩波書店.
- 松田映二(2006). 「郵送・インターネット比較調査で世論調査の可能性を探る」, 『朝日総研レポート AIR21』 No. 198, 190-219.
- 柳沢亨之・福田昌史(2015). 「メディア 英総選挙『接戦』大外れ 各世論調査会社」, 読売新聞, 2015年5月15日付朝刊.