

# 調査手法転換時の対応と判断

## —2000年総選挙と2001年参院選挙の事例—

松田 映二

### 1. はじめに

選挙に関する調査としては、投票日前に選挙の大勢を知るため有権者を対象に行う「選挙情勢調査」と、投票日当日の開票前に当落を知るため実際に投票した人を対象に行う「出口調査」の二つがよく知られている。ここでは、おもに2000年総選挙と2001年参院選の選挙情勢調査で朝日新聞社が取り組んだ新しい試みについて取り上げる。

一般的に、選挙情勢調査で議席を推計するためのプロセスは、①調査支持率から選挙得票率を推計する、②推計得票率を基に各候補者の当選の可能性を確率化する、③各政党別に当選確率を積み上げ推計議席とする、という流れになる。ただ、この大まかな三つの過程のそれぞれに気をつけねばならない重要なポイントがある。

過去データはあてになるのか——①の作業には、因果関係を探るための過去データが必要になる。しかし、参院選は3年に一度、総選挙は解散がなければ4年に一度のため、それぞれ約10年でようやく3回分のデータが蓄積できる。この3回の選挙戦の争点はそれぞれ違う。政党構図も分裂や新党発足などでがらりと様変わりしてしまうこともある。投票動向も様々である。無党派層が盛り上がり投票率が伸びなかったり、自民支持層が棄権する中で無党派層の投票行動が風を巻き起こすなど、過去データには一つとして同じ条件下のものはない。ゆえに砂浜から貝殻を拾うがごとき細かい個別の特徴は無視して大まかな共通の特徴をあぶりだすことが肝要である。

当選確率は正しいか——②の作業で、一般的には推計得票率は標本誤差などを考慮した正規分布と仮定すれば、各候補者の分布の重なり具合から当選確率を求められる[1]。しかし、小泉内閣発足前までは、自民

党候補は推計得票率からマイナスの方向に偏る傾向があり民主党候補はプラスに偏る傾向がある。つまり、分布が左右対称ではなく、各政党ごとに特徴があるのではないかという疑問が生じる。また、中選挙区制での公明党候補は、支持基盤から確実に票を集め、調査支持率が高なくても当選する可能性が高い。普通に分布の重なり具合から当選確率を求めれば、当選ライン近辺のライバル候補の分布と重なり当選確率は目減りする。ということは、公明党候補には通常の誤差理論から求めた誤差をつけずに、その誤差幅を思いっきり小さくすれば目減りを防げるのではないか。しかし、こうした各政党の個性を反映させるための手立てが裏目に出た場合は、推計は大きくはずれることを覚悟しなければならない。

議席予想は簡単か——③の作業では、②で求めた当選確率がそれほどはずれていなければ、たいした問題ではないようにみえる。しかし、ここが要注意なのである。ある選挙区の調査結果は、自民党のA候補が当選確率0.8、民主党のB候補は0.2だったとしよう。さらに別の方法で推計したとき、A候補の当選確率は0.7、B候補は0.3だったなら、どちらにしろこの選挙区ではA候補が当選しそうなのだからたいした問題ではない。ところが、こうした傾向が、衆院選の300小選挙区のうち1/3の100選挙区で起こったらどうなるだろうか。100選挙区で自民党候補と民主党候補の当選確率がそれぞれ例示のように逆方向に0.1だけずれるなら、トータルの獲得議席予想はそれぞれ $0.1 \times 100 = 10$ 議席ずれることになるのである。自民と民主の議席差は相乗効果で20議席動くことになる。300選挙区という数の多さは、推計担当者として脅威である。

こうした、技術的な問題のほかにも大きな問題がある。有権者が投票日近くにならないと投票態度を決めなくなったため、調査日時の設定が重要になったこと。また、衆院選の300小選挙区すべてを調査する場合、会場や人の手配などは一大プロジェクトとなること。

まつだ えいじ

朝日新聞社 世論調査部

〒104-8011 中央区築地 5-3-2

もはや面接調査は人手が不足して不可能である。このような理由で、調査方法の選択が、これまで以上に重要な問題となってきたのである。

## 2. 伏線

### 2.1 1996年総選挙での教訓

中選挙区制に代わり今の小選挙区比例代表並立制が初めて導入されたのは、1996年である。1選挙区に複数の候補者が当選するこれまでの中選挙区制の推計は、当選順位まで当てねばならず難しかった。そのため、すべて1人区になったことを歓迎したものである。しかし、実際は中選挙区制よりも小選挙区制の方が議席推計は難しかったのである。選挙区数が129から倍以上の300になったことや公示からの運動期間が14日間から12日間に短縮されたことが大きく影響している。

1996年はこれまでの面接調査から電話調査に切り替えた。電話なら調査員が半分で済むからである。電話回線を架設できる会場の手配などの新たな問題も加わり、東京(25選挙区)など多くの選挙区を抱える都市部は10月10日に「無風」、11日に「わりと無風」、12、13日に「激戦」と3群に分けて調査することになった。調査日数は「激戦」のみ2日間で無風とみられるものはほぼ1日間の調査になった。調査結果は散々だった[2]。選挙区だけでみると、自民183±9、

新進76±9の推計に対し、結果は自民169、新進96。とくに「無風」や「わりと無風」と読んで1日調査で早めに済ませた選挙区では自民69±5、新進29±5の推計に対し、結果は自民60、新進38と大きくはずしてしまった。これは定数1にもかかわらず鞍替えで現職が2、3人も立候補する選挙区があったり、「無風」選挙区の調査時点で投票態度がまだ決まっていなかったりしたことが原因と思われる。このとき得た教訓は、「調査はできるだけ投票日に近づけ2日間で実施すべし」というものだった。

### 2.2 1998年参院選での疑問

情勢調査結果をふまえた事前報道では「自民、過半数回復困難か」と見出しを打った[2]。過半数は届かないだろうが、そう大きくは落ち込まないだろうという意味合いだった。しかし、予想とは異なり、投票日翌日の7月13日付朝刊一面は「自民惨敗、首相退陣へ」の大見出し。自民党の議席は59±6、民主党は19±4とみていたが結果は自民44、民主27だったから、まさに情勢調査そのものも大惨敗だった。

事後分析をしてみたが、民主が躍進するだろうと思わせる兆候がデータのどこにも見出されなかった。「投票態度が調査日から投票日まで大きく動いたのではないか」という意見と「調査が有権者の意思をとらえ切れていなかったのではないか」という意見が激突し、その結論はいまだに出ていない。前者の理由の

1998年参院選東京選挙区  
連続トレンド調査(RDD)

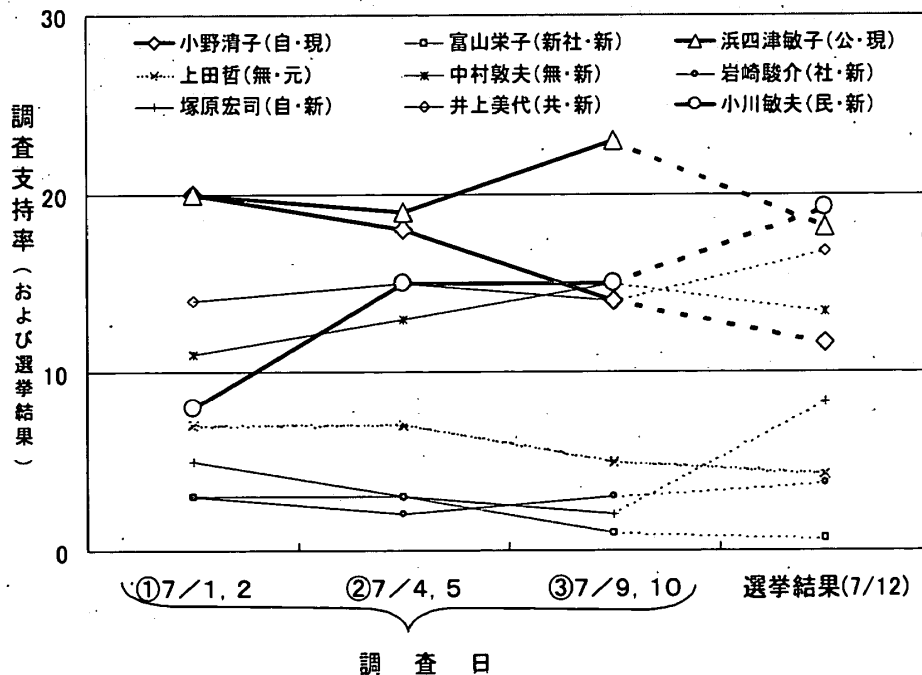


図1 1998年参院選東京選挙区の連続トレンド調査結果

一つは、東京選挙区のみで7月1、2日と7月4、5日と7月9、10日と3回連続のRDD (Random Digit Dialing) 法—コンピュータで電話番号をランダムに作成して電話するため電話番号非掲載者にも調査できる—によるトレンド調査を実施した結果が裏付けている。図1のように、民主党の小川敏夫氏の支持率が急増し、逆に自民党の小野清子氏は急落しているのがわかる。しかし、小川氏が公明党の浜四津敏子氏を抜いてトップ当選することまでは、このデータからは読み取れない。後者の理由の一つは、電話番号非掲載者に調査できない名簿方式の電話調査を採用したため、掲載率の低い都市部ほどその影響を受けたのではないかというものである。民主党躍進の大役を果たした無党派層ほど、実は電話番号非掲載者が多い。いずれにしろ、このときの検証では名簿方式の電話調査には限界があるのではという疑念を拭い去ることができなかった。

### 3. 2000年総選挙での試み

1996年総選挙の反省から、300小選挙区すべてを投票日1週間前の土日の2日間で同時に調査することにした。ただし、膨大な調査員集めや大量の電話回線を架設できる会場の確保などの作業負担が重く、すべて自前ではできないと判断した。また、1998年参院選の教訓から、テスト調査を進めて開発してきた新しい調査手法であるRDD法を導入することにした[3~6]。300選挙区のうち半分をこれまでの名簿方式電話調査で自前で実施し、残り半分をRDD法で外注することにした。すべてRDD法に切り替えなかったのは、調査方法が変わればデータの中身も変わり、過去データなしで推計式を作成しなければならなくなるからである。

名簿方式かRDD法かの選定は、まず県内の選挙区数が偶数の県と奇数の県の2群に分け、それぞれ県・選挙区コードの若い番号から並べる。次に、県・選挙区コードの若い方から2選挙区単位で乱数を発生させ、両者に振り分ける方法にした。その理由は、①各県の名簿方式の電話調査をできるだけ県内選挙区数の半分になるようにし、弊社支局担当者の負担に不公平が生じないようにすること、②300選挙区の縮図になるように半分の150選挙区を選ぶことで、過去データ(名簿方式)を生かして推計できること(推計結果を2倍する)、の二つがあった。さらに、大都市部でRDD法を多く採用すれば、新しい方式のデータが地域的に

表1 2000年総選挙における名簿方式とRDD法の推計比較表

		都市部		中間部		田舎部	
		選挙結果	推計議席	選挙結果	推計議席	選挙結果	推計議席
自民党	名簿方式	9	13	25	31	54	54
	RDD法	10	11	23	26	56	56
民主党	名簿方式	13	7	20	11	10	8
	RDD法	11	10	20	17	6	10

偏り比例区推計が難しくなるという判断もあった。

実際の推計でも多少工夫をした。中選挙区制の場合は当選定数別に推計式を用意していたこともあり、小選挙区制では定数1だが、都市部型、田舎型、そのどちらにも該当しない中間型と3分類して別々に推計式を作成した。表1は名簿方式とRDD法での各分類別の推計値と結果の比較表である。

都市部、中間部でRDD法の方がよく的中しているのがわかる。図2は調査支持率と選挙得票率の相関を表している。名簿方式は都市部ほどデータがバラつき45度線から大きく逸脱する。ところが、RDDでは都市部でも田舎部でも45度線からともに逸脱しない。

この形状に大きく影響しているものがある。それは候補者名を答えてくれた回答者数である。名簿方式のものでは、1000人の対象者のうち92人からしか候補者名を答えてもらっていない選挙区があった。当然ながら推計ははずれた。図3に示すように名あげ数の少ない選挙区ほどはずれていることがわかる。これら的大半は都市部の選挙区である。用意した1000サンプルのうち、電話番号非掲載など番号が判明しない対象者は脱落する。脱落が激しいところは調査可能な残りのサンプルが500程度に目減りすることもある。

さて、このときの情勢調査結果をふまえた事前報道では「自民、安定多数うかがう」と見出しを打ったが、投票日翌日の紙面は「自公保激減、民主躍進」の大見出しになり、またしてもはずれた[2]。このときの推計方針は、過去データのある名簿方式で実施した150選挙区の政党別推計議席数を2倍する、というものだった。ところが、調査直後、名簿方式の群とRDDの群での政党別議席数が大きく異なったため、急遽方針転換をした。初めて導入したRDDの結果も勘案して推計議席数を求めたのである。それでも、はずれてしまった。各県の第1区で民主党が健闘をみせる「1区効果」などが名簿方式の調査ではとらえきれなかったこと、そしてこれらの当選確率の読み誤りが積み積みもって全体として大きなずれになってしまったことが原因である。RDDのみを2倍していれば的中してい

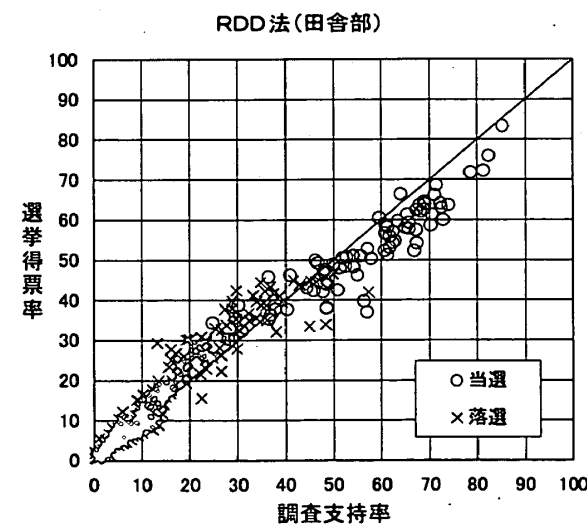
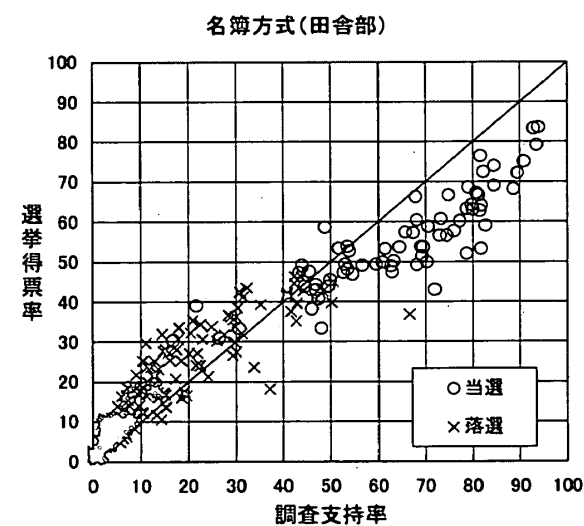
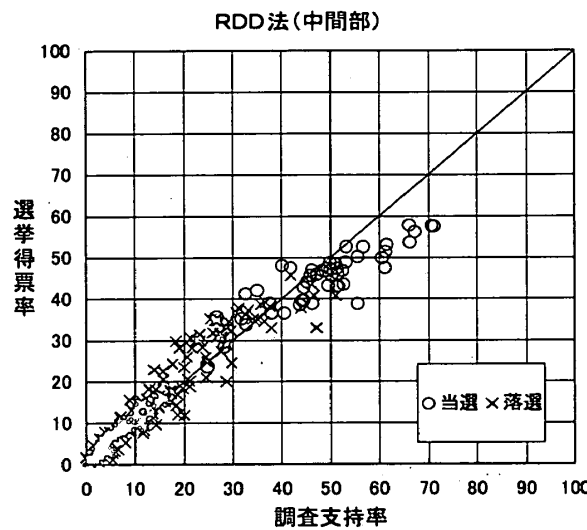
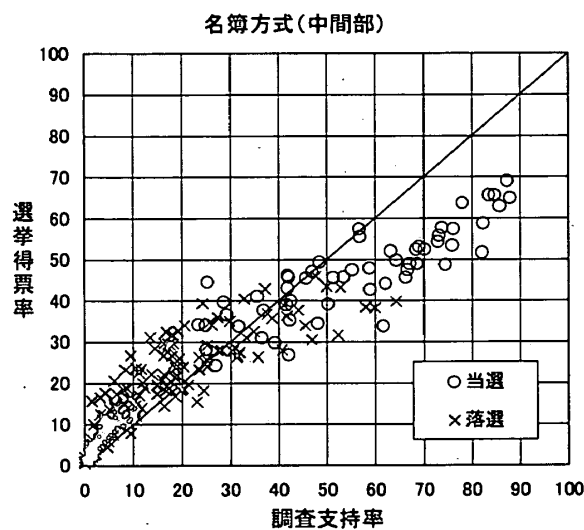
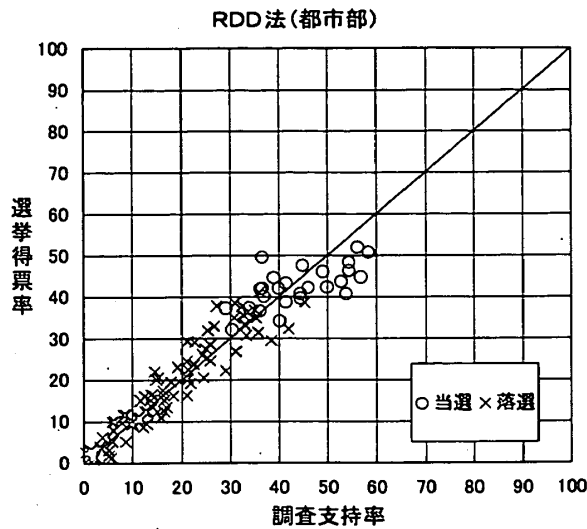
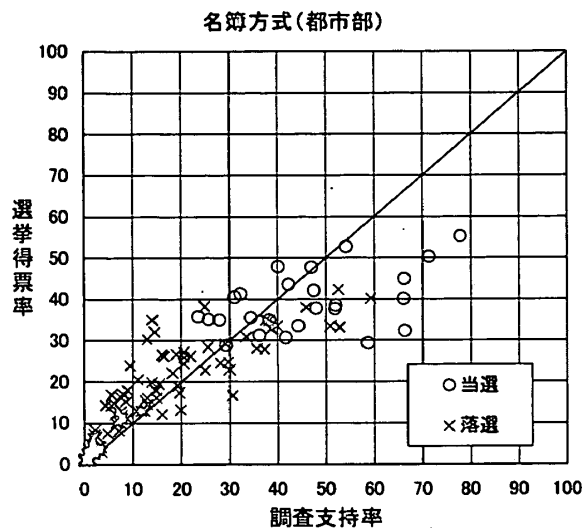


図2 2000年総選挙における支持率と得票率の相関—名簿方式とRDD法の比較—

たのだが、RDD群が正確な縮図になっているかどうか選挙が終わってみなければわからず、苦渋の選択だった。

「過去データはあてになるのか」と冒頭に述べたが、

調査手法に欠陥があれば過去データどころか当日のデータさえあてにならないということである。この時点での過去データには電話番号非掲載者の投票態度が反映されていないため、番号掲載者の投票態度から、掲

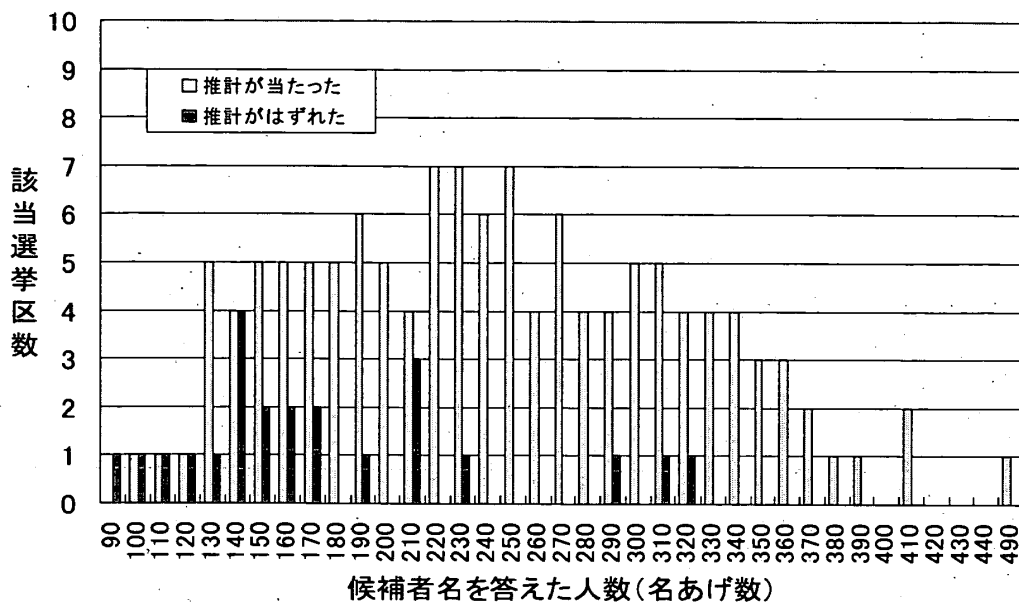


図3 名簿方式で実施した各選挙区の名あげ数と推計的中度の関係

載者と非掲載者を合わせた有権者（もっと正確に言えば投票者）の投票行動を予測しなければならない。番号非掲載者の行動が前回選挙と大きく異なったとき、推計式にはその行動を反映する手立ては何もないことになる。このことは、「対象者をできるだけカバーすることが調査の基本」ということを再認識させてくれる[7~9].

#### 4. 2001年参院選での試み

もはや調査はRDDしかないと確信して、全選挙区をRDDで実施した。前回参院選は名簿方式で実施したため、推計式作成のための過去データはなかった。しかし、推計式作成のためには、過去データは不可欠である。そこで、2000年総選挙で実施したRDD群の150選挙区のデータを基に推計式を作成することにした。

なにも簡単にこの方針を打ち出したわけではない。衆院選と参院選では有権者の投票行動に違いがあること、参院選では改選数が1~4と定数に幅があるが利用できるデータは小選挙区の定数1のデータしかないこと、が大きな懸念材料だった。しかし、名簿方式群の150選挙区の調査支持率と選挙得票率の相関図(図2)のうち、都市部データには参院選の改選数3~4の3次回帰曲線が一致し、中間部には改選数2、田舎部には改選数1の3次回帰曲線が一致した。参考までに確認しておくが、参院選での改選数3~4の選挙区は東京や大阪など大都市をもつ県である。改選数2は中都市をもつ県、改選数1は小都市をもつ県である。つ

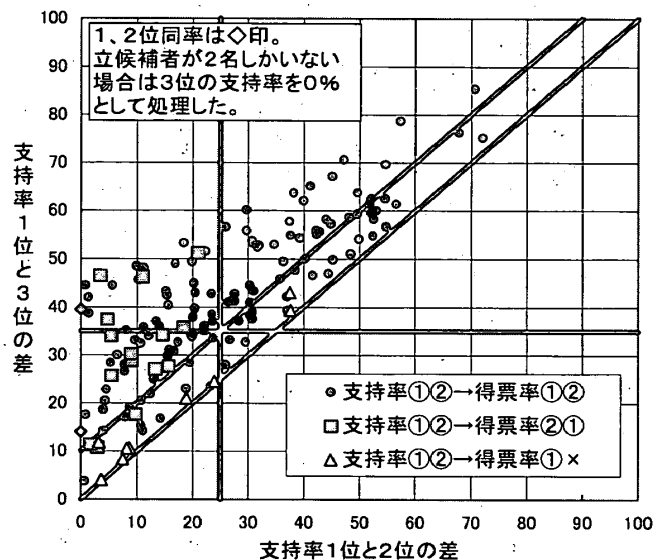


図4 RDD 150 選挙区の1, 2, 3位の対決構図

まり、同じ名簿方式の電話調査データから作成した回帰曲線だが、衆院選の当選定数1の都市部のデータ構造と参院選の改選数3~4のデータ構造が見事に一致したのである。つまり、データ構造はその選挙区の定数に依存するのではなく、都市規模に依存するということがわかったわけである。RDD群の150選挙区の調査支持率と選挙得票率の相関図(図2)をあらためて見直してほしい。電話番号非掲載者に調査できるRDDでは、各相関が都市規模で大きく変わらない。推察するに、衆院選の都市部のデータ構造と参院選の改選数3~4のデータ構造が似ているのは、ともに電話番号非掲載者が多い都市部だからである。掲載率が高い田舎部ではデータは45度線に近づくことでもわか

るように、データ構造は名簿方式という調査手法の欠陥によって変わっていたのである。ちなみに、都市部の回収率が極端に悪くなる面接調査のデータでも同様な傾向がみられる。

これで、RDD群のデータを利用した方がよいことはわかった。次は、どう利用するかである。そこで、150選挙区のデータから定数2の選挙区のデータとして利用できるものを選別することにした。対決構図を明確にするため、調査支持率順位1位と2位の候補者の支持率の差をX軸に、1位と3位の差をY軸にとると図4ができる。45度線は2位と3位の支持率が同じ場合を意味する。プロットのマークは、支持率が1位と2位の候補者が得票率でも同順位だったもの(●)、順位が入れ替わったもの(■)、支持率2位の候補者が得票率で3位以下だったもの(△)、の3種類に分かれる。■印の判別はX軸25%あたり、△印の判別は直線 $Y=X+10$ あたり、また■印の判別線に合わせてY軸の閾値を35%とした。参院選・改選数2の過去4回分のデータでプロットしてみると、 $X=25$ ,  $Y=X$ ,  $Y=35$ ,  $X=0$ で囲まれた台形領域にほぼ収まった。そこで、この台形領域にある選挙区のデータを用いて改選数2用の推計式を作成することにした。

さて、純粋な過去データがない中、どのような推計式を作成したのか。これまで朝日新聞社が用いてきた推計の手順は大まかに言って、①過去データの調査支持率(X)と選挙得票率(Y)から求めた3次回帰式 $[Y=aX^3+bX^2+cX+d]$ に今回調査の調査支持率を代入して仮の推計得票率を求める、②候補者が前職か新人かといった議員歴や候補者の年齢などから補正を加え推計得票率を確定する、③推計得票率は誤差を含むものなので各候補者の誤差幅の重なり具合から解析的に当選確率を計算する、④各候補者の所属政党ごとに計算した当選確率を積算して政党別獲得議席数とする、の4段階にわたる[12]。

この参院選のときは、過去データがないため正確な推計得票率を計算することができなかった。そのため、当選確率を推計得票率の誤差幅の重なり具合から求める方法ではなく別のものを考えなければならなかった。機密事項でもあり詳細に記述できないが、大まかな考え方を述べると、①数量化理論を使い「各候補者の支持率の差」や「候補者の議員歴」、「様々な階層からの支持模様」など各要因の項目ごとに得点を用意し、該当する項目に付けられた得点の合計値を候補者の強さとする、②候補者の強さと当選確率との相関から求め

表2 2001年参院選・選挙区の政党別推計議席数と選挙結果の比較

	選挙結果	推計議席
自 民	44	41 ~ 44 ~ 47
民 主	18	13 ~ 17 ~ 19
公 明	5	3 ~ 4 ~ 5
共 産	1	0 ~ 2 ~ 4
社 民	0	0 ~ 1 ~ 3
自 由	2	0 ~ 1 ~ 2
自由連合	0	0
諸 派	0	0
無所属	3	2 ~ 4 ~ 7
計	73	73

られる曲線に得点を代入して当選確率を求める、という手順になる[10~12]。これまでは、推計得票率を求めることが推計の要だったが、これをやめたわけである。こうした方式を採用したことの利点は、冒頭の「当選確率は正しいか」で述べたような純粋な誤差理論から求める当選確率が実態にそぐわないことを回避できたことである。

改選数3、4の推計式の作成は、不可能だった。小選挙区制では一つの選挙区で4人から6人の有力な候補者がいることはまずないからである。そこで、各候補者の支持率の差のみから当選確率を求めた。確率曲線は、参院選の過去4回分のデータ(名簿方式電話と面接)から各候補者の支持率の差と当選率の分布を求めて作成した。

情勢調査の結果は、「与党、過半数確保の勢い」と見出しが打たれ、ようやく的中した[2]。選挙区の推計議席と選挙結果を表2に示す。自民党の議席を的中させ、ほかの政党も±1のずれにとどまった。大転換をし、過去データのない状況下での推計作業となったが、ここまでのすべての考察が正しかったことが裏付けられたわけである。

## 5. おわりに

本稿では、朝日新聞社の選挙情勢調査への取り組みを可能な限り紹介し、RDD法という新しい調査手法が有効であるということを示した。残念ながら、選挙情勢調査は推計作業があるため、なかなか過去データの呪縛から逃れられない。そのため、同じ調査手法を続けることになってしまう。しかし、小選挙区制の導入で物理的に面接調査ができないこと、代替の名簿方式の電話調査で続けて推計がはずれたことなどが引き

金になり、大きく調査手法を転換することとなった。これでも大丈夫かという、そうでもない。これからも検討しなければならない問題は多い。二つ例示する。

〈1〉 RDD法は回答率の悪い調査である。電話をかけても不在で世帯用番号か法人番号かが区別できないものの総数を分母に加えて回答率を計算し直すと、4割程度に下がる。面接法のように7割には到底およばない。しかし、面接法でも、都市部ではオートロックの普及や深夜帰宅などで回収率は5割台に低下してきた。

〈2〉 携帯電話の普及が進み、家庭用電話が減少すれば、選挙情勢調査をRDD法で実施するのは困難になるだろう。家庭用番号は、市外局番、市内局番など電話番号そのものに地理情報が付加されているのに対し、携帯電話の番号にはついていない。小選挙区など狭いエリアでの調査でランダムに携帯電話番号を発生することは至難の業だろう。

新しい媒体の出現や生活様式の変化、選挙制度の改正など、今後も目的に合わせた調査手法の開発・改善が欠かせない。開発にあたって重視することは、①対象者すべてに調査可能か、②回答拒否者を減らせないか、③面接、電話、インターネットなど利用手段によるバイアスを把握する、の3点である。

調査関係者の方もそうでない方も、これまでの手法を大転換するときのデータ利用や考え方の参考になれば本望である。

## 参考文献

- [1] 林知己夫・高倉節子：「予測に関する実証的研究」, 統計数理研究所20周年記念号, 第12巻(第1号), pp. 9-86, 1964.
- [2] 朝日新聞：衆参院選挙情勢調査報道紙面, 1996.10.16, 1998.07.07, 2000.06.20, 2001.07.24.
- [3] 前田忠彦, 土屋隆裕：「日本人の国民性 2000年度吟味調査報告—電話・郵送・面接調査の比較—」, 2001.
- [4] 佐藤武嗣：社会調査ハンドブック, 「RDD調査」, 朝倉書店, 2002.
- [5] 松田映二：「findings：良質な調査結果を得るために (RDD法を採用した経緯)」, 新情報, Vol. 85, 2001.
- [6] 松田映二：「朝日新聞社のRDD調査について」, 行動計量学, 第29巻第1号(通巻56号), 2002.3.
- [7] Lepkowski, J. M. : Telephone Survey Methodology, "Telephone Sampling Methods in The United States", New York : Wiley, pp. 73-98, 1988.
- [8] Groves, R. M. and Couper, M. P. : Nonresponse in Household Interview Surveys New York : Wiley, 1998.
- [9] Couper, M. P. : "The Promises and Perils of Web Surveys", Association for Survey Computing, pp. 1-23, 2001.
- [10] 駒澤勉：「数量化理論とデータ処理」, 朝倉書店, 1982.
- [11] 駒澤勉：「パソコン数量化分析」, 朝倉書店, 1988.
- [12] 松田映二：「社会調査ハンドブック」, 「情勢調査」, 朝倉書店, 2002.