

選挙予測を巡る対話

日経リサーチ
理事
鈴木 督久

選挙予測調査と報道

主 今回は選挙調査の話だけれど、マーケティング・リサーチャーには興味ないかなあ・・・

客 予測調査という観点に立てば、製品開発・販売予測という場面には共通性もあるから調査会社にとっては選挙予測調査の経験が、市場調査においても何かの役に立つとは思うよ。

主 ちょうど 2003 年の衆院選が終わったので整理はしておきたいのだ。まだ分析が完全に終わっているわけではないけれど、途中報告のような格好で、どうかお許し頂きたい。

客 ところで選挙予測というと TV の「当確出ました！」という番組のイメージが世間では強い。

主 そうだね。あれは出口調査の結果を主に使っているけれど、俺の話は新聞社のやる約一週間前の情勢報道のための事前予測調査だ。

客 じゃあ、調査の前に予測報道の結果をまとめてくれよ。

主 トップ見出しは、日経・読売が「与党、安定多数の勢い」、毎日・共同・産経が「自民、単独過半数の勢い」、朝日は抑制して「自民、過半数うかがう」だっ

た。見出しに関する各社の考え方については解説しないが、紙面作りの現場では各社とも予測数字に忠実だ。

客 選挙結果は「民主、躍進」だったけれど、予測段階では、そこまで明言できなかったかな。

主 民主を高く予測していたのは朝日で「民主、大幅増の勢い」。他社は比例代表に着目して、読売は「民主、比例第1党も」、毎日が「民主、比例は優勢」、日経は「民主、比例で互角」という見出しだった。結果的に民主はいくつかの小選挙区でも接戦を制したね。

客 民主が予測段階よりも結果的に伸びた——という現象・傾向は前回 2000 年の衆院選と同じだ。トップ見出しも前回と、ほとんど同じだよ。

速報競争と調査の立場

主 紙面はそれくらいにして、調査に移ろう。表1に各社の調査概要をまとめた。今回の特徴の第一は、速報競争で掲載が1日前倒しになったことだ。産経以外の各社は日曜組だった。調査の立場からは、落ちていて日曜日の夜まで実査をやりたいと思う。編集の立場では他社より1日でも早く報道したい。少なくとも

も遅れるわけにはいかない。しかし、各社とも牽制しあって結局は同日掲載になってしまうのなら、そんな競争はやめて少しでも質の高い調査データを得る努

力をしよう——と調査の立場では言いたいのだがなあ。2004年の参院選はどうなることやら・・・

＜表1＞主要各社の調査概要

	掲載日	実施日	抽出法	計画標本	回収標本	回答率	未定	備考
日経	11/3	10/30-11/2	TDD	150000	69348	46	28	電話帳
共同	11/3	10/31-11/1	LSD	382,500	140000/259000	54	25	選挙人名簿
読売	11/3	11/1-11/2	LSD	225000	134162	60	3割	選挙人名簿
朝日	11/3	10/30-11/2	RDD		172650/274701	63	4割	番号空間
毎日	11/3	10/31-11/1	RDD		72936	*	4割	(割当法)
NHK	11/4	10/31-11/2	RDD		177636/244413	73	31	番号空間
産経	11/4	10/30-11/2	RDD		43385	*	35	(割当法)

RDD: Random Digit Dialing

LSD: List Sample Dialing

TDD: Telephone Directory Dialing

- ・ 「未定率」は小選挙区における態度未定者の比率(調査時点で未だ投票する候補者を決めていないと回答した人)
- ・ 毎日と産経は割当標本のため、回収率が定義されていない。

客 測定法は各社とも電話だが、標本抽出法に違いがあるね。

主 読売・共同・日経は電話帳掲載率の制約を受けるリスト抽出法だ。朝日とNHKはRDD。毎日と産経もRDDではあるが、割当法だと思われる。標本サイズは直接的に調査予算に関連するけれど、計画標本みて朝日・読売・共同・NHKはおおよそひとつの小選挙区あたり千人で設計している。日経と毎日はその半分だ。産経はさらに少ない。

客 標本サイズと標本抽出法の影響は、各社のデータを比較分析すれば、ある程度は推測できるが、現在はそのような共有環境にはないらしいね。

主 多額の費用をかけた調査データだから共有なんて開放性には到達できない。抽出法に関して、個人的な印象だけでいえばRDDの方が良い気がする。ただ、予測調査は世論調査と違って「よく予測する」ことが目的なので電話帳抽出を利用する判断もある。予

算の問題が抽出法を選択する要因として一番大きいけれどね。

客 標本サイズの影響はあったのかな。

主 難しい問題だけれど、大局的な趨勢を予測するという意味では、各社とも成功したといえる。しかし「自民の苦戦、民主の躍進」傾向をどこまで調査データに表現できたかという度合いでは、どうも標本サイズも関連がありそうな気がする。もちろん、原因は他にもあるさ。調査主体名、予測モデル、回収率など。いずれにせよ標本サイズは調査予算に直結しているからなあ。

2003年総選挙:小選挙区の予測

客 実際のデータではどうなんだ？

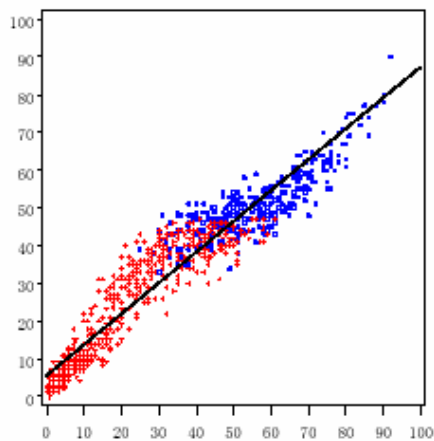
主 2003年の小選挙区について見てみよう。比例代表に関しては、いつか別の場所でやろう。図1は調査支持率と得票率との単回帰散布図だ。決定係数 92%

だからすごい説明力だろ. 他社の論文報告などを読んでも, 全体としてはこんな様子だ. 図2は予測得票率との単回帰散布図で, 解析によって決定係数を上昇させている. これが議席予測に使われる.

客 立派な予測ができそうな感じだね.

主 全体の趨勢を言い当てるのはいいけれど, 細部を検討すると議席予測は難しいよ. 300 選挙区のうち 200 議席は調査情報なしでも当選者が決まっているような選挙区だ. 調査支持率にも大きな差が出ている.

<図 1> 2003 年小選挙区: 調査支持率と得票率.
($n = 1026, R^2 = 0.92$)

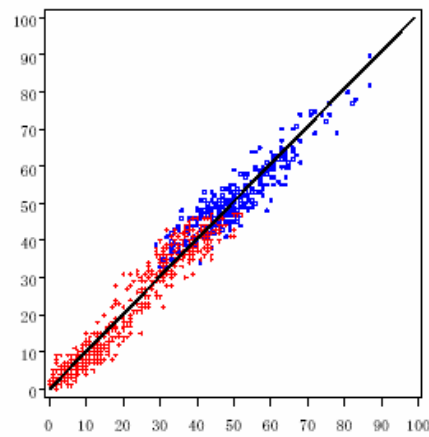


○: 当選, +: 落選

問題は接戦の選挙区だ. 決定係数が非常に高いけれど, それは実は「見かけの」良さだと言いたい. 調査支持率が 15%以下の候補者が当選することはないし, 65%以上で落選したこともない, というような経験知が各社ともあるだろう. 図1と図2は, そんな「両端」の候補者も含めた散布図なのだ.

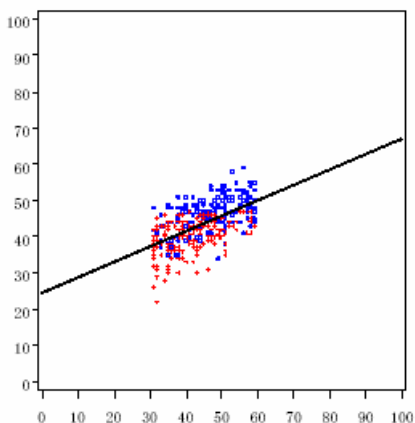
客 統計の教科書に出てくるような相関係数の話だね. ところで図3と図4はどういう意味だ?

<図 2> 2003 年小選挙区: 予測得票率と得票率.
($n=1026, R^2=0.96$)



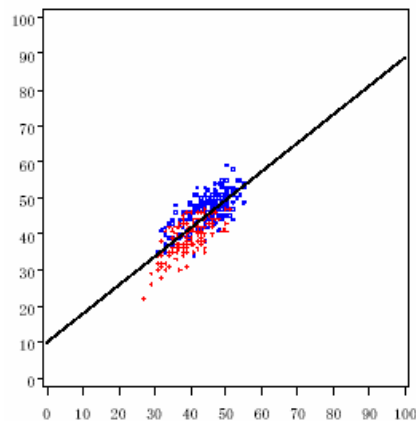
○: 当選, +: 落選

<図 3> 30% < 調査支持率 < 60%の候補者のみ.
($n = 352, R^2 = 0.35$)



○: 当選, +: 落選

<図 4> 30% < 調査支持率 < 60% の候補者のみ.
($n = 352, R^2 = 0.57$)



○: 当選, +: 落選

主 調査支持率が 30%~60%の候補者でデータを絞った結果だ。逆に言えば、落選する可能性の高い候補者と、当選する可能性の高い候補者を除外した残り。言い方は悪いが、分かりきったところを除いた結果だ。そして、ここが本当に予測分析の力量の発揮されるべき場所だ。ここは300人余りの候補者がほぼ半々で当落を決する領域で、ここでは調査支持率による決定係数は35%に過ぎない！ 予測得票率にしても57%だ。こうなると標本サイズの影響も気になる。各社のデータの品質に差が出ている部分かも知れない。こういう水準になると、調査データの質が悪ければ、いくら高度なデータ解析をしても壁にぶつかる。

客 君はいつも、「標本サイズが大きくて、そのデータが偏っていれば、間違いの方向に誤信を深めるだけだ」と言うじゃないか。ここでもそうかい？

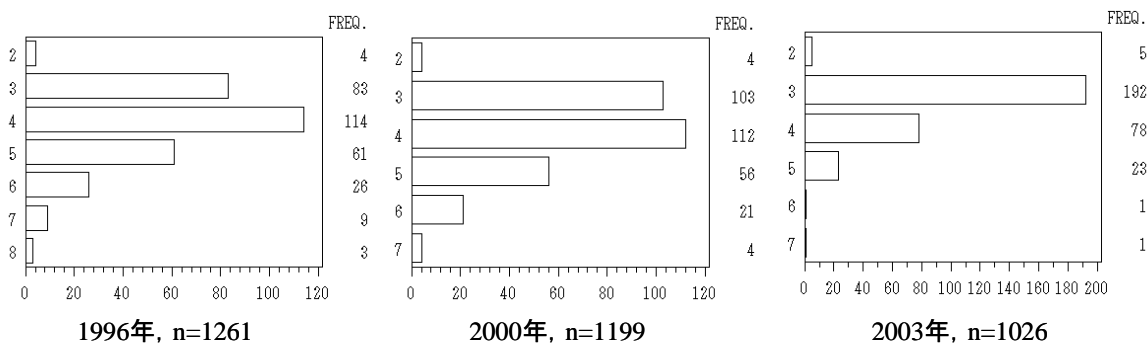
主 接戦の場合はデータ解析をしても追いつかない限界を感じることもある。統計的推測という理論的方法も無力だ。ただ「有意差なし」と言うのみ、広い信頼区間を示すのみだ。もっとも、激戦を的中させるのは

統計学の仕事ではないと思う。「何とも言えない」という統計学的表明は誠実な態度だ。データ解析は点推定値を洗練し、統計理論は区間推定するのが役回りだが、いずれにせよ高品質のデータこそが最後には勝つ。高品質性は妥当性よりも信頼性の確保において特に重要だ。だから調査フィールドの管理業務が永久に重要性を失わないのだ。

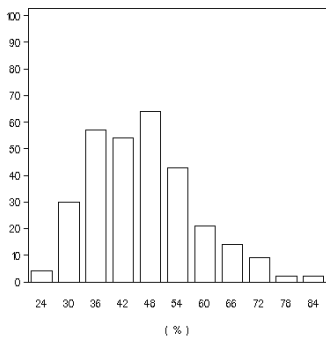
客 ところで今回の衆院選では二大政党制の傾向が強まったというが、予測との関連ではどうか。

主 立候補者が減ったね。小選挙区あたり1人分は減った(図5)。それに伴って得票率の分布にも変化が出ている。図6は当選者の得票率分布だが、50%前後での勝利に集中化している。一方、図7の落選者の得票率分布は大きく変化している。U字型に分布しているが、「一騎打ち」の接戦となる選挙区が増えているのだ。これが二大政党制化ということの選挙データによる表現だが、このような背景で予測モデルをどうするかは今後の課題だね。

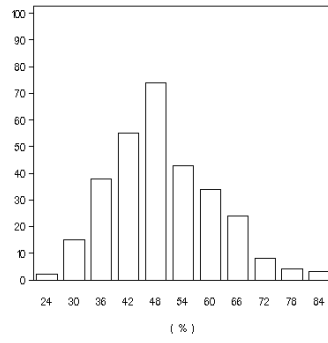
<図5> 候補者数の分布



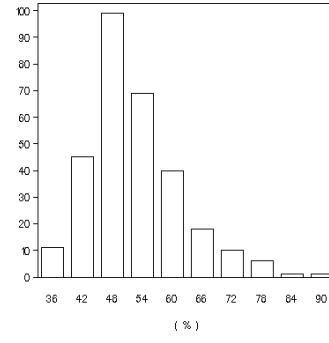
＜図6＞当選者の得票率分布



1996年, n=300

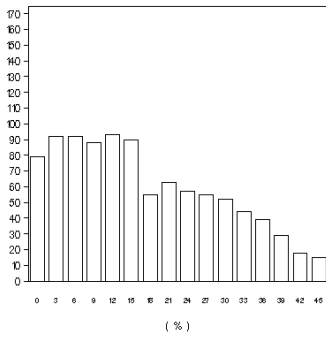


2000年, n=300

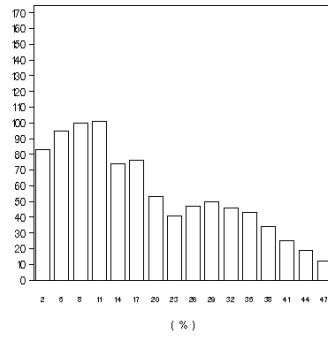


2003年, n=300

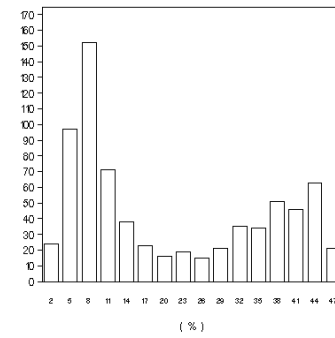
＜図7＞落選者の得票率分布



1996年, n=961



2000年, n=899



2003年, n=726