Mystery Pollster

Demystifying the Science and Art of Political Polling - By Mark Blumenthal

« About Those Tracking Surveys | Main | Likely Voters V: More Gallup »

October 27, 2004

LIKELY VOTERS IV - THE GALLUP MODEL

So how do pollsters select likely voters?

The best place to start is the Gallup likely voter model, the granddaddy of them all. Gallup is also worthy of special scrutiny for other reasons: It is easily the best-known brand name in survey research. Its campaign polls, conducted in partnership with CNN and USA Today, receive more attention and arguably have greater influence than other polls over campaign coverage. Finally, Gallup's methodology has also been the object of far more criticism this year than any of the others.

Before reviewing the Gallup model and its shortcomings, I want to strongly emphasize one point: **We are able to nitpick their model largely because Gallup has been extraordinarily open about their internal procedures, more so than other pollsters.** They have patiently answered questions from the most critical of outsiders. They routinely turn their raw data over to the Roper Center after each election, where academics can scrutinize their methods and search for flaws. That Gallup has been punished, in effect, for its openness has not been lost on competitors who remain considerably less forthcoming. So while it is appropriate to question Gallup's model, we ought to give them credit for their transparency. By opening themselves up to criticism this way, they are advancing the art and science of survey research.

Gallup has been open about its methods from the start. In 1960, Paul Perry, Gallup's president and research director, published an article in *Public Opinion Quarterly* detailing their election poll methodology ("Election Survey Procedures of the Gallup Poll," vol. 24, pp. 531-542). Then as now, respondents tended to over-report their true voting intentions, so selecting likely voters was not a matter of simply asking, "will you vote?" To identify the true "proportion of the population old enough to vote who will vote," Perry used internal validation studies that compared respondents' answers to their actual vote history. During the 1950s, Gallup sent its interviewers to vote registrar offices after each election to check whether their respondents had actually voted.

While no single question perfectly predicted whether a respondent would vote, Perry combined a series of questions "related to voting participation" into a 1-7 point scale that was highly predictive of actual turnout: "The system is such that the greater their likelihood of voting, the higher their score. Respondents are then ranked on the basis of their scores." Perry first set aside those who said they were not registered because their studies had shown that only "a negligible percentage of them vote, something on the order of between 1 and 5 percent." Then he used the index to select a subgroup of the highest

FAQ

Why are polls showing contradictory results?

What does the margin of error mean?

How do pollsters select "likely voters?"

Should pollsters weight by party identification?

Do undecided voters break for the challenger?

Are polls missing people with mobile phones?

EXIT POLL FAQ

Why "Mystery Pollster?

Other Questions

scoring respondents whose size matched the proportion of adults that typically voted in each election. In presidential and congressional elections from 1950 to 1958, the model reduced the average "deviation" from reality on Gallup's polls from 2.8 among registered voters to 1.1 percentage points among likely voters.

Although Gallup has made minor modifications, the questions and procedures that Perry described 44 years ago remain in use by the Gallup Poll today. Among those who say they are registered to vote (or who plan to do so before the election), Gallup uses the following questions to create a scale that varies from 0 to 7:

> 1) How much have you thought about the upcoming elections for president, quite a lot or only a little? (Quite a lot = 1 point)
>
> 2) Do you happen to know where people who live in your neighborhood go to vote? (Yes = 1 point)
>
> 3) Have you ever voted in your precinct or election district? (Yes = 1 point)
>
> 4) How often would you say you vote, always, nearly always, part of the time or seldom (Always or nearly always = 1 point)
>
> 5) Do you plan to vote in the presidential election this November? (Yes = 1 point)
>
> 6) In the last presidential election, did you vote for Al Gore or George Bush, or did things come up to keep you from voting?" (Voted = 1 point)
>
> 7) If "1" represents someone who will definitely not vote and "10" represents someone who definitely will vote, where on this scale would you place yourself? (Currently 7-10 = 1, according to this "quiz" on *USA Today*)

A few additional notes: They automatically exclude from the likely voter pool anyone who says they do not plan to vote (on #5). They also give anyone 18-24 an extra point, to help make up for having said they did not vote in the last election (perceptive readers will immediately sense a problem here -- I'll take that up in the next post).

According to Gallup's David Moore, they aim this year to select a pool of likely voters equal to 55% of their adult sample - their estimate of the appropriate "turnout ratio" likely in this election. In practice, the percentage that scores a perfect 7 out of 7 typically comes very close to 55%. If it ever goes over, they will tighten the scoring of the last question about likelihood to vote (giving a point to those who answer 8-10, for example, instead of 7-10), so that likely voters will always be some combination of sixes and sevens this year.

The one hitch is that they usually have more than enough sixes to bring the total size of the likely voter pool to 55%. So Gallup weights down the sixes to make the weighted value of the likely voters equal to 55%. An example makes this easier to follow: (although the following numbers are *totally hypothetical - I made them up*): Suppose the pool of those scoring 7 out of 7 is 50%, and the sixes are 10%. They would then weight down the value of the sixes by half (multiply times 0.5): 50% + (10%*0.5) = 55%.

What if the sevens are 50% and the sixes are 15%? They would weight the sixes by 0.33: 50% + (15%*0.33)=55%. Make any sense?

Important concept: Gallup does not claim that this model perfectly predicts who will vote, only that the pool of likely voters consists of those *most likely to vote*. They also designate some voters as likely and others as not likely. In these two respects, their model is consistent with virtually other pollster. From there, however the way

RECENT POSTS

The NYT Reader's Guide to Polls

Vacation

Crosstabs.org

So What *Is* A Push Poll?

Another Online Poll on Online Activity

CT: Party ID from Quinnipiac Poll

More on the CT Exit Poll Experiment

Incumbent Rule Redux

New Polls from Fox, AP

Connecticut Epilogue

CATEGORIES

Debates

Divergent Polls

Exit Polls

FAQ

Focus Groups

General

Hillary Clinton

Immigration

Impeachment

Incumbent Rule

Initiative and Referenda

Innovations in Polling

Instant Reaction Polls

Internet Polls

Interpreting Polls

Iraq

IVR Polls

JunkPolls

Likely Voters

Measurement Issues

Microtargeting

Miscellanous

Mobile phones

MP Housekeeping

Polling & the Blogosphere

Polls in the News

Pollsters

President Bush

pollsters pick likely voters diverges in a big way.

In the next post, the shortcomings and critiques of the Gallup model..

[See other discussions of Gallup's seven question model by the *Wall Street Journal*, *Salon.com* and Ruy Teixeira]

Related Entries - Likely Voters

Posted by Mark Blumenthal on October 27, 2004 at 01:02 PM in Likely Voters | Permalink

Comments

I've been swayed by your arguments that weighting by party ID doesn't make much sense. Lately, though, some liberal sites have been posting internals (I think mostly from Gallup, though maybe Wapo/ABC also) showing some likely voter screens leading to very strange demographic results compared to past election turnout (eg undersampling minorities). How can pollsters justify the extraordinary lengths they go to to achieve a representative sample of registered voters, and then throw this out when it comes to the LV results?

I apologize if this foreshadows your own critique.

Posted by: cw | Oct 27, 2004 3:11:48 PM

I also have the same question as cw re: Gallup weighting by party affiliation and a recent Florida poll, but will eagerly await your next installment. Excellent article.

Posted by: Anthony Stevens | Oct 27, 2004 4:09:10 PM

Speaking of Gallup, which naturally brings up the topic of cell phone usage and the roll that plays in surveys, here is an interesting article on the demographics and behaviors of cell phone users:

http://www.wired.com/news/politics/0,1283,65473,00.html?tw=wn_tophead_4

Posted by: Observer | Oct 27, 2004 5:11:51 PM

I notice your say that Gallup gives an extra point to 18-24 year olds. What about newly naturalized citizens ? Or are these too small a subgroup to count ? I think they certainly make a difference in local elections (say in CA or FL or TX).

Also, what about people who've already voted ? Are they counted as part of likely voters by default ?

Posted by: erg | Oct 27, 2004 6:58:32 PM

Excellent find Mark!

According to Gallup I am not a "likely voter". I vote religiously and yet by their qualifications I am not a "likely voter". The reason is I moved this year (same state) and my polling place/precinct has changed, I have never voted there before and I do not yet know where the actual location will be. All of my other answers show a highly motivated person commited to voting who has always been that way... and yet somehow I'm not a "likely voter"!

Might explain the rash of extreme/outlying Gallup polls...

Posted by: Brent | Oct 27, 2004 9:48:44 PM

So according to Gallup I am not a "likely voter" despite being a well-

educated, spending hours a day reading about the election (including this blog!), and having voted in every presidential election I've been eligible to vote in. I moved this year, so have never voted here and haven't managed to track down my polling place yet. I see the post above says the same thing. So are we looking at a t.o much greater than 55% or is this a bias in who is answering the phone at night?

Posted by: Jason | Oct 28, 2004 1:24:44 AM

Any one individual could take their personal situation and induce that it is applicable to a greater portion of the population than it truly is.

A personal example: I score a 7 but missed the 1998 election because my car broke down while I was travelling out-of-state and wasn't drivable until I missed the election.

The key point is that Gallup needs to determine a set of questions that will give them a pre-determined "likely" voter turnout percentage that comports with the historical turnout results. It is a black box solution.

Posted by: Eric | Oct 28, 2004 8:06:43 AM

FORGET THE POLLS. BUSH WINS.

Polling has been historically unreliable predictor of the presidential elections. However, three other polls were predictable. And they all point to Bush victory.

1. Iowa Electronics Market shows Bush victory. It was wrong only twice.

2. Readers Weekly shows Bush victory. It was always rights since 1956.

3. Halloween Masks Sales shows Bush victory. It was alwasy right.

Will all three predictors be wrong this year? It is possible. Everthing is possible.

But can you say with 98% accuracy that they will be right, as they have so consistently been right? Absolutely.

Posted by: Hermes1LA | Oct 28, 2004 10:03:32 AM

Gallup should probably be more flexable. In 1992 turnout was 72% and this year it looks to be much higher as well, even possibly as high as 80%. Rigidly adhering to 55% seems odd. Also not weighing by interest of different demographic groups seems odd. Blacks really came out for Gore and made it a close election in 2000. As the country gets more diversified turnout amongst one demographic group could easily change an election. A lot have changed since 1960.

Posted by: jocko | Oct 28, 2004 10:33:24 AM

Hermes, I sincerely hope that was satire. It's Weekly Reader, by the way, not Readers Weekly.

Add to your list the Nickelodeon poll that has never been wrong since it began in 1988 and that gave the nod to Kerry.

And the consumer confidence "poll" where, since 1968, no sitting president has ever won when consumer confidence is below 100 (as it now is).

There are others, of course, but all of these things are nonsense.

ElectionProjection.com

Electoral-Vote.com

R. Chung's Graphics

Prof. Wang's State Poll Meta-Analysis

Prof. Pollkatz Pool of Polls

Rubicon Electoral College Update

Wall Street Journal Poll Graphic

Slate: Election Scorecard

The Polling Report

ACADEMIC SURVEYS

General Social Survey (GSS)

The Roper Center (Survey Archives)

National Annenberg Election Study (NAES)

Univ. of Wisc./BYU 2004 Election Study

Joint Program in Survey Methodology (JPSM)

Institute for Social Research (ISR)

National Election Studies

SURVEY RESEARCH ORGS

The Council for Marketing and Opinion Research (CMOR)

The World Association for Public Opinion Research (WAPOR)

Council of American Survey Research Organizations (CASRO)

The National Council on Public Polls (NCPP)

American Association for Public Opinion Research (AAPOR)

POLITICAL COMMENTARY

Kausfiles

Andrew Sullivan

Instapundit

Daily Kos

Posted by: PaulB | Oct 28, 2004 10:33:36 AM

I may be missing something. But isn't there a problem with entirely excluding unlikely voters. We know that some unlikely voters will vote and we know that some likely voters will not vote. If one of these groups two favors one candidate more than the other (which we can assume to be true, since this is the point of likely voter models), the results will be distorted. Wouldn't it make more sense to _weight_ for likelihood-to-vote?

Posted by: PeterS | Oct 28, 2004 10:54:46 AM

Of course, 18-24 year olds will NEVER make it into the six and seven point respondents in the top 55% 'likely voter' model, because despite being given an extra point for their age, they lose 3 points on questions 3, 4, and 6 as a result of not having been eligible to vote before. So if this is your first opportunity to vote in 2004, you maximum possible 'likely voter' score will be five. You will never be included. That and the party affiliations explain a lot about Gallup. Thanks, MP.

Posted by: Paul | Oct 28, 2004 10:54:50 AM

I understand the rationale for not weighting by party ID or party registration: both can always change from year to year.

Why don't polls ask who the person voted for last time and then weight the sample to make it reflect the national popular vote in 2000? If we kept getting surveys where only 45% of people voted for Gore, we could correct for that bias.

Posted by: Ari | Oct 28, 2004 1:28:03 PM

Mark: I love your site and the detail, and just noticed my link. Thanks!

And thanks for making the point about averaging polls! Makes me feel all warm inside.

Posted by: winston | Oct 28, 2004 5:27:49 PM

I just wanted to make one last comment to this thread. There is one site by a professor who uses statistics to predict a presidential election. Professor Ray C Fair from Yale has set up an equation that takes into account a bunch of different variables and calculates the percentage of how much a candidate will get vote wise.

here is the link:
http://fairmodel.econ.yale.edu/vote2004/index2.htm

To see the latest prediction click on the October 29, 2004 prediction. If you want to know how he does it and see his predictions all the way back to 1916 in using data, click on the top link on the page - November 2002 update paper.....

Even if you take away the average error rate over the last 88 years, which is 1.495% up to a maximum of 5% which is the 1992 election in which they didn't account for Ross Perot, Bush still wins by a comfortable margin.

There was a study done by midlevel folks in the federal economics department that basically came up with the same spread of 52-57% win for Bush. I think they used about the same calculations.

Oh and about the Weekly Reader and Nick polls, lets add one more. The OneVote poll (link: http://channelone.com/election_2004/results/)which is a high school poll of 1.4 million teens shows a 55% win for Bush which I

think is close to what the Weekly Reader came up with. I wouldn't take
much stock into the Nickelodeon poll.

Posted by: Michael | Oct 31, 2004 9:05:41 PM

The comments to this entry are closed.