Notes on the Freeman/Mitofsky American Statistical Association - Philadelphia presentations

Friday, October 14, 2005

1:15 pm Polling Bias or Election Fraud? –
Steve Freeman (University of Pennsylvania)

2:00 pm 2004 Exit Polls: What Bloggers And Others Got Wrong –
Warren Mitofsky (Mitofsky International)

Summary: The talk was not as well attended as we expected, probably about 60 people in all. Mitofsky brought a table's worth of people including Lenski. I probably had about 25 people there, and another half a dozen with whom I'd communicated by email but never met.

The Center for Organizational Dynamics has a publicist who tried to get TV and print coverage. Neither CSPAN nor other TV stations came, but one of the most prominent writers on polls and polling, Richard Morin of the Washington Post, did come and I did speak with him briefly. A documentary maker came to film the event, but Mitofsky refused to provide them with a release. I only know of one news story published. I'll send it, and my response, in another message.

Thanks to the excellent suggestions by Josh Mitteldorf, Stephanie Singer, Jim Murphy, Dava Guerin, Tom Guggino and Vince Salandria – I was able to prepare excellent text and slides. [available here: http://www.appliedresearch.us/sf/epdiscrep.htm]

Here's what Richard Morin wrote to a mutual contact:

> The debate was interesting. Warren was much more subdued than I expected (a bit too subdued, I thought), and Steve was a bit more polite than he was in your rehersal. …

> If anything, Steve had too much to say. I thought Warren delivered a bit too much of his exit poll stump speech and should have done more to respond to Steve's specific points. But a very good debate and perfect for my profile of Warren.

[I have no idea what he is talking about when he said that I was "more polite than .. in your rehersal."]

I do not have Mitofsky's talk, but I think Morin characterizes it fairly. He calmly and confidently described how they make their Election Day estimates. He addressed none of the 17 improbabilities and neglected correlations that comprised the central body of my presentation. Josh pointed out that he seemed to be making our point for us that he conducts the polls very professionally, and there is good reason for confidence in the original estimates and the data.

Only towards the very end he addressed my talk, and then simply that he didn't know where I came up with my numbers. He concluded with a slide that he claimed "kills the fraud argument," although I don't think anyone in the room quite understood why it would do so. He didn't even identify the axes.

Rebuttal: I made a few clarifications after his presentation:

(1) All the exit poll data I used in my study came from Edison/Mitofsky, most of it from the 77 page Inauguration eve report. For most of my analyses, I used Precinct Level Disparity (PLD), which is the difference between whom people said they voted for as they walked out of the voting booth, and the way those votes were officially recorded.

(2) Be careful with pollster-speak. Not only WPE/PLD; but "correcting" the exit poll data means adjusting it to conform to the count; and "Democratic Overstatement" may well be, and certainly to at least some degree would more accurately be called ""Democratic Undercount."

(3) He made a big deal of "selection interval" explaining the discrepancy, but I pointed out that "selection interval" may well be a spurious correlation, in that it likely co-varies with actual predictors, notably "paper" and probably "rural"

(4) Likewise for "co-located precincts" which allows for the Ballot rotation mischief that has been documented in Cleveland, and "interviewer distance from polling place." Ken Blackwell likely had good reason why he didn't want observers close to the polling place.

(5) It's hard to know what to make of any chart of residuals. For example, does it include outliers? Even the Albert Einsteins of the world have to show methods and data.

(6) The data that has been made available is not the data that is useful for important analyses. Everyone knew what data was important:

  (a) Precinct level summary data to could physically investigate precincts with high PLD, some of which have impossibly high PLD: Why were they so high?

  (b) County identifiers to statistically examine effects of voting technology and partisan control.

But instead, Edison/Mitofsky released individual level data and now claim they cannot permit inspection of such data because it conceivably could compromise respondent confidentiality by linking it to this individual level data.

Of course, even the most private, personal census data – which this isn't – is studied, as is the most sensitive national security data (which this, in fact, may be). Using procedures such as clearances, background checks, and on-site inspection. Since shortly after the election I offered to work on-site at Mitofsky International along with whatever team of statisticians and social scientists they chose, and to raise funds to reimburse Mitofsky International for whatever costs were incurred from our conduct of this research. He said he could not permit even this.

(7) When they speak about "improvements" for next year, note that they will not permit any release of "uncorrected" exit poll data. Be careful about such an "improvement." Is that something we consider desirable? To never again receive exit poll information that can shed perspective on official election results?

Some Lessons:

(1) Try to be crystal clear on the data, in particular, the importance of PLD: Where it comes from. Why it is such an important measure. What it means.

(2) Do not let them dismiss anything. At the point, I was explaining my use of PLD, Lenski said from the back of the room that you can't do that, "that's not how we used it." I said I know that but that I was using the data to clarify the discrepancy, and he threw up his arms and snorted dismissively. I don't think I let him get away with it, but next time, I'd be even tougher, and put him on the spot until he either acknowledges the validity of the point or his willful ignorance is completely apparent to the audience.

(3) Several people came away confused about "early results." Add to the list of misleading pollster-speak that when Lenski or Mitofsky speaks of "early exit poll data," or the "half-time score" that although the implication is of early-in-the-day results, they actually mean end-of-day data that is not yet adjusted to conform to the count.

(4) Continue to hammer on the proprietary data. In fact, we really should press more on this individual confidentially argument. Not all pollsters support Mitofsky's position, but I haven't gotten the details on that debate.

(5) If they indicate that they know best because they are the authorities – and that only they have the data, point out not only that it's a breach of scientific ethics that they make these claims without sharing the data, but they have a vested interest in seeing the data as they do.