Polling Phollies

George Rebane
4 September 2016

No one can escape the polls during this election season whose results will be deemed cataclysmic by at least half of the American electorate, no matter who wins. So all kinds of people from government, campaigns, and academia are attempting to peer into the collective public mind to determine its sentiment and portents. The instrument of choice for this endeavor is the poll. A poll is a type of survey which consists of applying a questionnaire to a sample of respondents drawn randomly from the population of interest.

Polls are abundant, there being at least twenty of them prominently quoted on a daily basis. To make sense of the plethora of divergent results, desperate people attempt to amalgamate them by various kinds of averaging each of which can stand about as much scrutiny as the legislature making laws, or worse, a bureaucratic brew drafting regulations. Here I want to disabuse the reader of giving much credence to any of the, say, ten polls to which he is subjected on any given day. The poll's result is usually given by a set of percentages attached to alternative answers, percentages that are supposed to sum to 100% so that the waterfront is appropriately covered.

To add credibility to the poll, the media also often include the 'margin of error' and the number of people questioned or the sample size. As an article of faith the poll's consumer is supposed to believe that the population was properly sampled and that the questionnaire was not loaded with agenda-driven queries meticulously wordsmithed. For the present exercise let's all join that body of the faithful and concentrate only on the claimed credibility bestowed by stated sample size and margin of error.

We can dispense with sample size since most people instinctively understand that the larger the sample, the more reliable should be the results. The understood measure of reliability is the proximity of the sample proportion/fraction/percentage to its unknown true value in the population, and that is communicated by giving the +/- interval around the reported fraction that would contain the real value, say, 95% of the time. What does 95% of the time mean? It means that if we were to repeat the survey with that sample size a gazillion times and record the result values, then 95% of them would fall within the stated +/- interval. (Techies will recognize this as the approximate four sigma region around the mean of the set of gazillion survey results.)

But for a multi-part answer that has three or more alternative choices – say the current candidates of the four parties who will compete for the presidency – that cited 95% confidence interval does not hold for all alternatives. Such a single confidence interval will work only for a question limited to two stated choices which exhaust the possibilities. Then its confidence interval may be computed from theory via a very complex equation the results of which are plotted in the famous 'banana curves' familiar to those of us who have designed and analyzed surveys.[1]

[1] Crow E.L., et al, *Statistics Manual*, Dover, 1960, p270ff. It is remarkable that such handy curves relating proportions and confidence intervals to sample size are not readily available on line, but the entire book can be downloaded as a pdf of images here - http://www.dtic.mil/dtic/tr/fulltext/u2/149334.pdf

To examine and illustrate this situation in more detail I wrote a little piece of code to simulate polling results for a four attribute distribution for various size samples. I chose the four presidential candidates' preference ratings as the attributes assigning the 'true values' or percentages to be 2%, 7%, 44%, 47% to approximate the current standings for Stein, Johnson, Trump, and Clinton respectively. We look at the results of five polls having sample sizes of 100, 300, 500, 800, and 1,000 respondents. The sample means or reported polling results for the various sample sizes are given in the table below (bold italicized numbers).

	SampleMeans				
TrueFrac	100	300	500	800	1000
2.00%	***2.03%***	***2.00%***	***2.00%***	***1.99%***	***2.00%***
7.00%	***7.03%***	***7.04%***	***6.99%***	***7.02%***	***7.02%***
44.00%	***43.84%***	***43.94%***	***44.08%***	***43.99%***	***44.05%***
47.00%	***47.09%***	***47.02%***	***46.93%***	***46.99%***	***46.92%***

From the simulations we calculate the +/- 95% intervals and show them below in the corresponding spaces for each candidate's true fraction and the poll's sample size.

	+/- 95% (2-sigma) confidence intervals				
TrueFrac	100	300	500	800	1000
2.00%	***2.88%***	***1.57%***	***1.22%***	***1.00%***	***0.87%***
7.00%	***4.93%***	***2.91%***	***2.30%***	***1.84%***	***1.57%***
44.00%	***9.99%***	***5.75%***	***4.41%***	***3.46%***	***3.14%***
47.00%	***10.13%***	***5.72%***	***4.51%***	***3.46%***	***3.11%***

The same data from the above two tables is perhaps more meaningful if presented graphically as in the figure below. There we see the simulation results plotted along the attribute fraction or x-axis, and the sample sizes vertically on the y-axis. Looking at the lowest row of sample size 100 results indicates that these are pretty much meaningless due to the size and extent of overlap of their confidence intervals. As we look at the results of polls with higher sample sizes, the situation improves, but not so much for the 44% and 47% attribute fractions. While the intervals for Stein (2%) and Johnson (7%) separate as we increase sample size, there remains considerable overlap for Trump (44%) and Clinton (47%). The amount of overlap indicates the likelihood that the actual percentages for the overlapping candidates can actually be reversed – e.g. that Trump actually leads Clinton by, say, 46% to 45%. And that significant possibility still holds when we sample 1,000 people from the voter population.

As the attribute fractions get close to the 50% or 0.50 levels, confidence intervals also reach their largest values. To reliably separate the confidence bounds on candidates whose approvals are in this range requires a large polling sample of, say, two to three thousand respondents. Such polls are seldom conducted on populations whose preferences are very dynamic such as those with candidates in tight races. Samples reaching up to about 1,000 respondents is the usual maximum, and that, as we see, still gives results with large overlapping confidence intervals. It is for this reason that historically last minute polls have been dramatically wrong when the real

election results come in. And the added takeaway from understanding these results is that the celebration of election victories from early poll results is totally unwarranted.