

[Institute of Mathematical Statistics](#)

Fostering the development and dissemination of the theory and applications of statistics and probability

[Renew / Join IMS](#)

Opinion Polling: Its Secret Sauce is also its Spoilage Source

December 15, 2020

The XL-Files

Xiao-Li Meng writes:

On November 6, 2020, I woke up to a flood (for a statistician) of tweets about my 2018 article, [“Statistical Paradises and Paradoxes in Big Data \(I\): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election”](#). A kind soul had offered it as an explanation to the question: “What’s wrong with polls?”, which led to the article going viral.

As much as I was flattered by the attention, I was disappointed that no one had asked “Why would anyone expect polls to be right in the first place?” A poll typically samples a few hundred or thousand people, but it aims to learn about a population many times larger. For predicting the US presidential election, conducting a poll of size $n=5,000$ to learn about the opinions of $N=230$ million (eligible) voters is the same as asking just about 2 people out of every 100,000 voters on average. Isn’t it absurd to expect to learn anything reliably about so many from opinions of so few?

Indeed when Anders Kiær, the founder of Statistics Norway, proposed the idea to replace a national census by “representative samples” during the 1895 World Congress of the International Statistical Institute (ISI), the reactions “were violent and Kiær’s proposals were refused almost unanimously!” as [noted](#) by former ISI President Jean-Louis Bodin. It took nearly half a century for the idea to gain general acceptance.

The statistical theory for polling might be hard to digest for many, but the general idea of representative sampling is much more palatable. In [a newspaper story](#) about Gallup Poll going to Canada (*Ottawa Citizen*, Nov 27, 1941), Gregory Clark wrote,

“When a cook wants to taste the soup to see how it is coming, he doesn’t have to drink the whole boilerful. Nor does he take a spoonful off the top, then a bit from the middle, and some from the bottom. He stirs the whole cauldron thoroughly. Then stirs it some more. And then he tastes it. That is how the Gallup Poll works.”

The “secret sauce” for polling, therefore, is thorough stirring. Once a soup is stirred thoroughly, any part of it becomes *representative* of the entire soup. And *that* makes it possible to sample a spoonful or two to assess reliably the flavor and texture of the soup, regardless of the size of its container. Polling achieves this “thorough stirring” via random sampling, which creates, statistically speaking, a miniature that mimics the population.

But this secret sauce is also the source of spoilage. My 2018 article shows how to mathematically quantify the lack of thorough stirring, and demonstrates how a seemingly minor violation of thorough stirring can cause astonishingly large damage due to the [“Law of Large Populations” \(LLP\)](#). It also reveals that the polling error is the product of three indexes: data quality, data quantity, and problem difficulty.

To understand these terms intuitively, let’s continue to enjoy soup. The flavoring of a soup containing only salt would be much easier to discern than a Chinese soup with five spices. *Problem difficulty* measures the complexity of the soup, regardless of how we stir it or the spoon size. *Data quantity* index captures the spoon

size, relative to the size of the cooking container. This shift of emphasis from only the sample size n to the sample fraction n/N , which depends critically on the population size N , is the key to LLP.

The most critical index and also the hardest one to assess is the *data quality*, a measure of the lack of thorough stirring. Imagine some spice clumps did not dissolve completely in the cooking, and if they have more chance of getting caught by the cook's spoon, then what the cook tastes is likely to be spicier than the soup actually is. For polling, if people who prefer candidate B over A are more (or less) likely to provide their opinions, than the polling will over- (or under-) predict the vote shares for B. This tendency can be measured by the Pearson correlation — let's denote it by r — between preferring B and responding (honestly) to the poll. The higher the value of $|r|$ (the magnitude of r), the larger the polling error. A positive r indicates overestimation, and a negative r underestimation.

The whole idea of stirring thoroughly or random sampling is to ensure r is negligible, or technically, to ensure it is on the order of the reciprocal of the square-root of N . Statistically, this is as small as it can be since we have to allow some sampling randomness. For example, for $N=230$ million, $|r|$ should be less than 1 out of 15,000. However, for the 2016 election polling, r was -0.005 , or about 1 out of 200 in magnitude for predicting Trump's vote shares, as estimated in my article (based on polls carried out by YouGov). Whereas a half a percent correlation seems tiny, its impact is magnified greatly when multiplied by the square-root of N .

As an illustration of this impact, my article calculated how much statistical accuracy was reduced by $|r|=0.005$. Opinions from 2.3 million responses (about 1% of the eligible voting population in 2016) with $|r|=0.005$ has the same expected polling error as that resulting from 400 responses in a genuinely random sample. This is a 99.98% reduction of the actual sample sizes, an astonishing loss by any standard. A quality poll of size 400 still can deliver reliable predictions, but no (qualified) campaign manager would stop campaigning because a poll of size 400 predicts winning. But they may (and indeed some did) stop when the winning prediction is from 2.3 million responses, which amount to 2,300 polls and each with 1,000 responses.

What was generally overlooked in 2016, and unfortunately again in 2020 (but see [this Harvard Data Science Review article](#)), is the devastating impact of LLP. Statistical sampling errors tend to balance out when we increase the sample size, but systematic selection bias only solidifies when sample size increases. Worse, the selection bias is magnified by the population size: the larger the population, the larger the magnification. That is the essence of LLP.

When a particular bit of soup finds itself on the cook's spoon, it cannot say, "Well, I'm a bit too salty for the cook, so let me jump off this spoon!" But in an opinion poll, there is nothing to stop someone from opting out because of the fear of the (perceived) consequences of revealing a particular answer. Until our society knows how to remove such fear, or the pollsters can routinely and reliably adjust for such selective responses, we can all be wiser citizens of the digital age by always taking polling results with a healthy grain of salt.

Originally published by *Scientific American*, December 6, 2020: <https://www.scientificamerican.com/article/the-secret-sauce-in-opinion-polling-can-also-be-a-source-of-spoilage/>

Institute of Mathematical Statistics
toll free (US): 877.557.4674
tel: 216.295.2340
fax: 216.295.5661
email: ims@imstat.org
© 2022 Institute of Mathematical Statistics

Website by [Sunray Computer](#)
[✉ Contact Us](#)



[Privacy Policy](#) | [Site Map](#)