November 1, 2018

News You Can Use

WHAT’S COMING UP?

We have a number of webinars planned to help you with using StatPREP materials or to better understand some of the technological tools available to help you run a more data-centric course. Several previous webinars are archived on our website and can be viewed here. We also have more live webinars coming up, so keep your eyes out for the next one!

WHO’S WHO?

Are you looking for specialized help, but don’t know who to contact? Check out our comprehensive list below that shows “who’s who” on the StatPREP Leadership team!

◊ Need help with R Studio? Contact Danny Kaplan (kaplan@macalester.edu)

◊ Need help with Little Apps? Contact Danny Kaplan (kaplan@macalester.edu)

◊ Need help finding a data set? Contact Danny Kaplan (kaplan@macalester.edu)

◊ Need help incorporating StatPREP materials into a lesson? Contact Kate Kozak (kathryn.kozak@coconino.edu)

◊ Need help assessing a StatPREP lesson? Contact Ambika Silva (Ambika.silva@canyons.edu)

◊ Have a question about the StatPREP website? Contact Jenna Carpenter (carpenter@campbell.edu)

◊ Want to know more about StatPREP webinars? Contact Kate Kozak (kathryn.kozak@coconino.edu)

◊ Want to know more about an upcoming StatPREP Workshop? Contact Mike Brilleslyper (mike.brilleslyper@usafa.edu)
A REVIEW OF THE PROCESS OF POLLING

In 1996 my department chair handed me the first statistics textbook I had ever seen. That single gesture constituted my college’s faculty development program for teaching statistics. One of the earliest examples in the book was about the importance of random sampling. It included a picture of President Truman holding up the Chicago Tribune’s infamous “Dewey Defeats Truman” headline. It’s a good story, but hardly timely, having taken place 48 years earlier. Few of my students knew who Truman was and none of them knew anything about Dewey.

Our students have grown up in an era of “scientific” polling. Being scientific, the results are reported with a margin of error, often ±3 percentage points, to help us know when conclusions are warranted and when not. Many of our statistics courses feature units on constructing a margin of error on a sample proportion, often with explicit reference to political polls. But, like Dewey defeating Truman, the story is no longer timely. The “error” in the “margin of error” is now only a small part of the unreliability of polls. Why?


For those of you reading this after the polling ends, I’ll describe the action. As I write this, 2,070,469 telephone calls have been made. In each Congressional district, the results from the past calls are laid out in a long line of circles, filled red or blue depending on the respondent’s response. But only 1 or 2% of the dots are filled. The large majority are empty: no response. Each new call generates a wiggling box at the head of the line of dots. It wiggles until the end of the call. Almost always, the box turns into an unfilled circle.

The poll I’m watching now, New Jersey 3rd district, is in its early stage. 4250 calls producing 62 responses. The margin of error? There’s a simple but meaningful statement laid right on top of the grayed-out tally so far: “Don’t take this poll seriously until we reach at least 250 people. We’re at 62.”

The calls are made based on a random selection from the phone numbers known to be in the district. But the random selection hardly generates a random sample when the response rate is 2%. To get something that resembles the population, pollsters weight their results. The New York Times is weighting “by age, party registration, gender, likelihood of voting, race, education and region, mainly using data from voting records files compiled by L2, a nonpartisan voter file vendor.” And then there’s the “likely voter” model, an informed guess about what fraction of people in each weighting strata will actually vote. There’s a detailed explanation in this article on the site, where the faulty results from the 2016 presidential election are attributed to a failure to weight by education level.

Seeing the polling process in such detail reveals our misconceptions about what’s important in statistics. The so-called “margin of error” is not an adequate indicator of the reliability of the poll. Instead, we need to be thinking about the factors used in weighting and the extent to which they capture the current configuration of political schisms. Polls are now about big, multivariable data (the “voting records compiled by L2”) and building models of turnout based on previous elections.