How many voters are needed to predict the election? The shift from convenience to random sample polls.

Unit 4 Lecture 1

Jonathan Auerbach

STAT 490 Pre-Cap Prof Development

jauerba@gmu.edu

November 2, 2021

How can scientists tell that a sample represents a population?

Is two million responses enough to predict an election?

▶ The Literary Digest, a weekly magazine with circulation over one million in the 1930s, conducted a poll to predict the winner of the 1936 election.

▶ The magazine mailed ten million questionnaires, "drawn from every telephone book in the United States, from the rosters of clubs and associations, from city directories, lists of registered voters, classified mail-order and occupational data."

▶ There were 2,376,523 responses, suggesting Kansas Governor Alfred Landon (R) would beat incumbent Franklin Delano Roosevelt (D) by 54 percent of the popular vote (370 electoral votes).

▶ How accurate was this estimate?

Sample questionnaire sent by the literary digest (Lohr and Brick 2017)

Standard error negligible under random sampling

- Let X be the number voting for Landon, where $X \sim \text{Binomial}(n = 2,376,523,\ p)$.
  - ▷ The Literary Digest observed $x = 1,283,322$.

- An estimate for p is $\hat{p} = \frac{x}{n} = \frac{1,283,322}{2,376,523} = .54$.

- An approximate 95-percent confidence interval for p is
  $$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .54 \pm 2\sqrt{\frac{.54(1-.54)}{2,376,523}} = .54 \pm .0006$$

```
binom.test(x = 1283322, n = 2376523)$conf.int
```

```
## [1] 0.5393659 0.5406336
## attr(,"conf.level")
## [1] 0.95
```

But FDR won by 60% (523 electoral votes)

What went wrong?

▶ The Literary Digest accurately estimated the proportion of respondents voting for Landon NOT the proportion of voters in America.

▶ Selection Bias/Incomplete Sampling Frame: Less than half of households owned a phone in 1936.
   ▷ Richer residents were more likely to be mailed a ballot, and these residents were more likely to vote Republican.

▶ Nonresponse Bias: Only one fourth of those contacted responded.
   ▷ The survey occurred in the midst of the Great Depression. It is again more likely that richer residents would respond, and these residents are more likely to vote Republican.

Meanwhile, Gallup conducted a (more) random sample of 50,000 voters

▶ Gallup actually had higher nonresponse rates, but took steps to avoid bias, including in-person interviews. Twenty-three thousand indicated they would vote for Landon.

▶ An estimate for p is $\hat{p} = \frac{x}{n} = \frac{23,000}{50,000} = .46$.

▶ An approximate 95-percent confidence interval for p is
$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .46 \pm 2\sqrt{\frac{.46(1-.46)}{50,000}} = .46 \pm .004$

```
binom.test(x = 23000, n = 50000)$conf.int
```

```
## [1] 0.4556239 0.4643808
## attr(,"conf.level")
## [1] 0.95
```

Gallup's standard error was seven times larger...

- ▶ ...but randomization reduced selection/nonresponse bias, giving Gallup the correct prediction of the winner (although also underestimating the popular vote).

- ▶ Gallup went even further: he released his survey before the Literary Digest and correctly predicted how far off the Digest would be.

- ▶ These bold but accurate predictions gave Gallup's American Institute of Public Opinion national recognition and lead to a massive improvement in polling statistics.

Pittsburgh Post Gazette Poll Summary (11/2/1936)

Straw Vote Fight Arouses Interest

Literary Digest and American Institute Are Far Apart
In Pre-election Forecast—Roosevelt, Landon
Both Get Around 56 Per Cent

With the American Institute of Public Opinion forecasting the re-election of President Roosevelt in tomorrow's voting, and The Literary Digest showing a victory for Governor Landon, the battle of the pre-election polls commands almost as much public interest as the election itself.

The Digest, after sending out more than 10 million ballots, found Governor Landon polling 57 per cent of the major party vote, and leading in 32 states, with 370 electoral votes.

The Institute of Public Opinion, operating on a sampling method, which calls for a representative cross-section of voters in every state and in all walks of life, predicts Mr. Roosevelt's re-election with approximately 56 per cent of the major party vote, and shows him leading in 40 states, of which 31 with 315 electoral votes are called "sure," and the others too close for positive prediction.

While not indorsing the results of either poll, The Pittsburgh Press here presents a state-by-state comparison of the final major party vote in both polls. For convenience, the Digest figures have been reduced to percentages.

Pittsburgh Post Gazette Poll Summary Continued

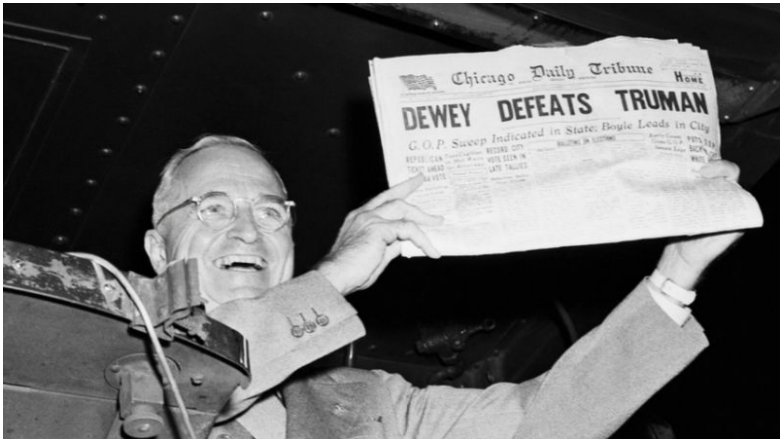| State	Electoral Vote	DIGEST % Major Party Vote	INSTITUTE -% Major Party Vote
Alabama	11	(Dem.) 77%	(Dem.) 83%
Arizona	3	(Rep.) 54%	(Dem.) 58%
Arkansas	9	(Dem.) 74%	(Dem.) 81%
California	22	(Dem.) 54%	(Dem.) 59%
Colorado	6	(Rep.) 61%	(Dem.) 55%
Connecticut	8	(Rep.) 68%	50-50*
Delaware	3	(Rep.) 59%	(Dem.) 53%
Florida	7	(Dem.) 58%	(Dem.) 69%
Georgia	12	(Dem.) 77%	(Dem.) 84%
Idaho	4	(Rep.) 58%	(Dem.) 61%
Illinois	29	(Rep) 61%	(Dem.) 51%*
Indiana	14	(Rep.) 62%	(Dem.) 52%*
Iowa	11	(Rep.) 63%	(Dem.) 51%*
Kansas	9	(Rep.) 64%	(Rep.) 52%*
Kentucky	11	(Dem.) 55%	(Dem.) 60%
Louisiana	10	(Dem.) 68%	(Dem.) 83%
Maine	5	(Rep.) 69%	(Rep.) 62%
Maryland	8	(Dem.) 51%	(Dem.) 60%
Massachusetts	17	(Rep.) 77%	(Rep.) 52%*
Michigan	19	(Rep.) 67%	(Dem.) 51%*
Minnesota	11	(Rep.) 60%	(Dem.) 54%
Mississippi	9	(Dem.) 88%	(Dem.) 95%
Missouri	15	(Rep.) 57%	(Dem.) 57%
Montana	4	(Rep.) 56%	(Dem.) 61%
Nebraska	7	(Rep.) 61%	(Dem.) 53%
Nevada	3	(Rep.) 51%	(Dem.) 67%
New Hampshire	4	(Rep.) 77%	(Rep.) 56%
New Jersey	16	(Rep.) 68%	(Dem.) 51%*
New Mexico	3	(Dem.) 51%	(Dem.) 60%
New York	47	(Rep.) 54%	(Dem.) 54%
North Carolina	13	(Dem.) 72%	(Dem.) 70%
North Dakota	4	(Rep.) 54%	(Dem.) 64%
Ohio	26	(Rep.) 61%	(Dem.) 52%*
Oklahoma	11	(Dem.) 51%	(Dem.) 64%
Oregon	5	(Rep.) 52%	(Dem.) 61%
Pennsylvania	36	(Rep.) 60%	(Dem.) 51%*
Rhode Island	4	(Rep.) 75%	50-50*
South Carolina	8	(Dem.) 85%	(Dem.) 94%
South Dakota	4	(Rep.) 65%	(Rep.) 52%*
Tennessee	11	(Dem.) 67%	(Dem.) 70%
Texas	23	(Dem.) 71%	(Dem.) 82%
Utah	4	(Dem.) 57%	(Dem.) 65%
Vermont	3	(Rep.) 74%	(Rep.) 61%
Virginia	11	(Dem.) 62%	(Dem.) 68%
Washington	8	(Rep.) 58%	(Dem.) 61%
West Virginia	8	(Rep.) 57%	(Dem.) 52%*
Wisconsin	12	(Rep.) 62%	(Dem.) 57%
Wyoming	3	(Rep.) 62%	(Dem.) 52%*

*NOTE—Because last-minute changes might cause a shift of one or two per cent in the vote, the Institute does not predict how the states marked (*) will go. The vote there as shown in its final poll varies from two points below 50 per cent to two points above 50 per cent.

SUMMARY OF VOTE

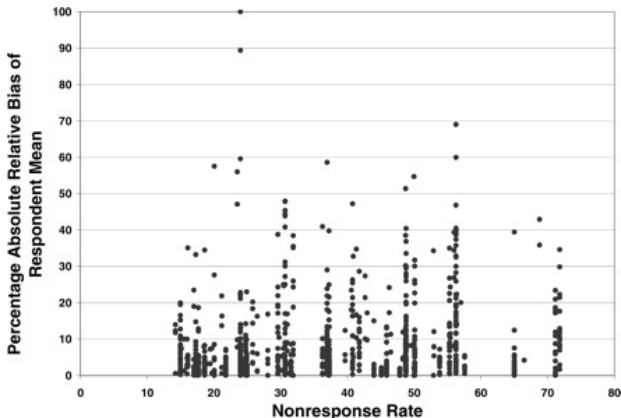|	Digest (Final)	Institute (Final)
Winner	Landon	Roosevelt
Per cent Major Party Vote	56% Landon	55.7% Roosevelt

Election prediction is far from perfect...

▶ Gallup incorrectly predicted New York Governor Thomas Dewey (R) would beat incumbent president Harry Truman (D).

▶ The incorrect prediction was famously printed by newspapers that were too impatient to wait for the full returns.

Are low-response surveys automatically biased?

▶ Groves and Peytcheva examine fifty-nine methodological studies, designed to estimate the magnitude of nonresponse bias.

▶ Not a strong relationship between the response rate and bias, although results sensitive to how bias is calculated.

References

1. Bryson, Maurice C. "The Literary Digest poll: Making of a statistical myth." The American Statistician 30.4 (1976): 184-185.

2. Gallup, George. "Opinion polling in a democracy." Statistics: A Guide to the Unknown (1972): 146-152.

3. Groves, Robert M., and Emilia Peytcheva. "The impact of nonresponse rates on nonresponse bias: a meta-analysis." Public opinion quarterly 72.2 (2008): 167-189.

4. Lohr, Sharon L., and J. Michael Brick. "Roosevelt predicted to win: Revisiting the 1936 Literary Digest poll." Statistics, Politics and Policy 8.1 (2017): 65-84.