Election polls

For each race, we have one or more polls reported as $\bar{X} \pm 2\mathrm{SE}$, or "$\bar{X}$ with a margin of error of ME." Typical ME's are 0.03-0.04. In our notation, $\mathrm{ME} = 2\mathrm{SE}$.

The data underlying $\bar{X}$ are $X_1, \ldots, X_n$, which are binary variables where $X_i = 1$ corresponds to support for one candidate and $X_i = 0$ corresponds to support for the other candidate.

If the poll is a simple random sample,

$$\mathrm{ME} = 2 \cdot \mathrm{SE} = 2\sqrt{EX \cdot (1 - EX)/n}$$

Solve for n:

$$n = 4EX \cdot (1 - EX)/\mathrm{ME}^2.$$

If $EX = 1/2$, and $\mathrm{ME} = 0.03$, then $n \approx 1100$.

Bias in election polls

Ideally, EX is the result that will occur in the actual election (since the election is a poll with huge sample size). In reality, it's more complicated for at least three reasons:

- People may change their mind between the time that the poll is conducted and the time of the election.
- People may not respond honestly in the poll.
- The people who respond to the poll may not be representative of the people who vote in the election.

Errors in election polls

The poll error is

$$\epsilon = \bar{X} - \mu,$$

where μ is the true support level for a candidate (ideally, $\mu = EX$). We can treat ϵ as being normally distributed with mean $EX - \mu$ (the poll bias) and standard deviation $\mathrm{ME}/2$. We can also write

$$\epsilon = (\bar{X} - EX) + (EX - \mu),$$

where $\bar{X} - EX$ is the random error in the poll, and $\eta \equiv EX - \mu$ is the systematic error in the poll.

Combining polls for the same race

Suppose we have several polls for the same race, giving us $\bar{X}^{(1)}$, $\bar{X}^{(2)}$, etc. We might average them to get a better estimate of μ:

$$\tilde{X} = (\bar{X}^{(1)} + \cdots + \bar{X}^{(m)})/m$$

If the polls are independent, the ME of $\tilde{X}$ is

$$\sqrt{(\mathrm{ME}_1^2 + \cdots + \mathrm{ME}_m^2)}/m,$$

and the bias is the average of the poll biases

$$(E\bar{X}^{(1)} - \mu + \cdots + E\bar{X}^{(m)} - \mu)/m$$

Combining polls for the same race

For example, combining two independent polls with $\text{ME} = 0.03$ gives an estimate with $\text{ME} = \sqrt{0.03^2 + 0.03^2}/2 = 0.02$.

However, it is very unlikely that the polls are independent.

It is also rather unlikely that the biases cancel substantially.

Forecasting aggregate Senate results

Suppose we are interested in forecasting which party will control the Senate.

Out of 100 Senate seats, 37 are up for election this year. Of the remaining 63 seats, 40 are held by Democrats and 23 are held by Republicans. Thus, the control of the Senate will shift to the Republicans if they win at least 28 of the 37 contested seats.

The Republican candidate is favored (most recent poll) in 27 of these races.

So the polls (taken at face value) favor the Democrats in the Senate. Greater uncertainty benefits the Republicans.

Forecasting aggregate Senate results

Suppose we have individual polls $\bar{X}^{(1)}$, ..., $\bar{X}^{(37)}$ for the contested seats, all with the same margin of error (for simplicity).

Then we could run a simulated election by generating election results

$$\bar{X}^{(1)} + E_1, \ldots, \bar{X}^{(37)} + E_{37}$$

where the E_j are normally distributed random values with mean zero and standard deviation ME/2.

Forecasting aggregate results

Then for each simulated election we can check whether at least 28 of the contested seats were won by Republicans. We can do this many times to estimate how likely it is that the Republican party will gain control of the Senate.

What if the errors are correlated?

We can generate correlated errors by applying problem set 5, #3d:

$$E_j = \mathrm{SE}_j \cdot (\sqrt{1-r}A_j + \sqrt{r}U)$$

where the A_j and U are independent, standard normal values.

Forecasting aggregate House results

There are 435 House seats, all representatives are up for election.

There are 110 landslide elections, of which Republicans will win 36.

Republicans must win 182 of the 325 non-landslide elections to take control of the House. The Republican or independent candidate is favored in 195 of these.

The polls (taken at face value) favor the Republicans in the House. Greater uncertainty benefits the Democrats.