The Survey

Observations

- Very low response rate $<10\%$.
- 20% is considered low on email surveys. 30% is more the norm.
- Response rate is in line from Lynelle's slides.
- No news on if students have been contacted about needs.

Fielding Notes

- Fairly ineffective email with buried lede.
- One reminder out. This is low, even the quick political polls will do two reminders.

Analysis Issues

- Self-selection/Participation Bias
- Open response questions are very rich.

Why is Selection Bias an Issue?

- Early survey respondents, and late respondents, tend to be different than those that respond later.
- They are motivated, either in a positive or negative sense.
- Highest participation rate was big survey on energy use in California during the crisis, 2000-2001, more than 50% phone with open-ended questions.

Example Bias

Survey about cookies

- Two groups Likes Surveys and Hates Surveys.
  - Likes responds 90% of the time
  - Hates responds 10% of the time
  - Equal sizes
- But liking cookies and surveys is correlated
  - 75% of the likes group also likes cookies
  - 25% of the hates group likes cookies

True State of the World

What fraction of the population likes cookies?

$$.5(.75) + .5(.25) = 0.5$$

This is fraction of population times rate at which they like cookies.

But you gave a survey

$$.9(.75) + .1(.25) = 0.7$$

This is probability of responding times the rate they like cookies.

Post Sampling Stratification

This is when we can identify the groups from other characteristics. In our case, they wear shirts that say likes and hates surveys.

IRL

- We do this on race/ethnicity and income.
- Generally, low and high income low response rate
- Non-Hispanic Caucasians have the highest response rate.

Post Sampling Stratification

We take observable characteristics that are correlated with response rate and use those to make the sample look more like the population.

- Those that are more likely to respond count as less than 1 observation.
- Those that are less likely to respond count as more than 1 observation.

$$Weight = \frac{Population\ Share}{Sample\ Share}$$

In Our Cookie Survey

$$Weight_{hate\ survey} = \frac{.5}{.1} = 5$$

$$Weight_{like\ survey} = \frac{.5}{.9} = 0.5555556$$

Note that the denominator is the probability they respond. In general weights are proportional to the inverse of the response rate.

What Happens

$$\frac{.5}{.9}.9(.75) + \frac{.5}{.1}.1(.25) =$$
$$.5(.75) + .5(.25) = 0.5$$

It gets us back to the known population number.

What Makes This Hard IRL

- Depends on observable characteristics
  - If you have little out of sample data you are stuck.
  - Pro Tip
- High amounts of uncertainty about the probability of taking survey conditional on observable characteristics is BAD.
  - Weak instruments in IV
  - Usually F>10 or F>15
  - Note lots of P-Hacking. Meta analysis shows big spikes at 10 in published empirical work.
  - Low F, you have bias still but you had it before. Could be less bias, could be more
- Explaining why to clients

In Our Survey