

# Data Science

Ethics

July 14, 2021

# Goals for today

Get you thinking about the ethics of doing Data Science

1. This is not about *morals*, lots of definitions of these things, for us:
  - 1.1 *Morals*: how you, individually, determine what's right or wrong
  - 1.2 *Ethics*: moral principals that govern a person's behavior *when conducting an activity in a professional capacity*
2. I can't tell you what's right or wrong, that's up to your moral worldview
3. I can tell you what the Data Science community is discussing as part of their ethical framework

# Statistical Inference

We've concerned ourselves with building models

1. Models are used to go from world of 'data' back to something we can change about the world
2. If we don't intend on our models being actionable, why have them?
3. Things we may conclude from a model:
  - 3.1 An estimate
  - 3.2 rejection of a hypothesis
  - 3.3 Clustering/classification of data points into groups

# Sampling

We often can't know all the data out there, so we have to *sample*

1. Usually, we want a random sample
2. When might we not want a random sample?
3. When gather data, (e.g. sampling the population), we have to be careful
  - 3.1 How were questions worded?
  - 3.2 How did you poll people?
  - 3.3 How is missing data handled?

# Potential Sources of Bias

## 1. Sample Bias

- 1.1 Selection Bias: some subjects more likely to be selected
- 1.2 Volunteer Bias: people who volunteer are not representative
- 1.3 Nonresponse Bias: people who decline to be interviewed

## 2. Survey/Response Bias

- 2.1 Interviewer Bias
- 2.2 Acquiescence Bias: tendency to agree with all questions
- 2.3 Social Desirability Bias: Reluctance to admit to embarrassing things

## 3. Confirmation Bias

- 4. Anchor Bias: you might say yes to something because a worse alternative was shown first.

# Potential Sources of Bias

How do clinical trials work (often see as the gold standard)

1. Some receive treatment, others in control group
2. Each group is picked completely at random
3. Considerations
  - 3.1 Only ethical of alternatives have a good basis (i.e. we don't ask folks to *start* smoking for a trial)
  - 3.2 Very expensive
  - 3.3 Sometimes impossible!
4. What about how people 'opt in'?
5. Social networks run trials on us all the time, is that okay?

# Final Thoughts

1. No easy answers
2. Play, explore, think
3. Use off-the-shelf technologies wherever possible
4. Think about possible introduction of biases and be skeptical of 'clear' results

# Course Evals

1. If you're able, do the course evals
2. I try to take feedback seriously (though feedback doesn't always agree!)





Thanks for your time!

Have a good summer!