Final Thoughts on Experiments and Causality

1. Observation vs. Experiment
   1. Data analysis allows for making decisions. Decisions involve counterfactuals. Existing in the state of the world where one has done X or Y.
      1. Observational data: compare units with different X values.
         1. Attempts to Fix Observational Data. All modeling choices will make tradeoffs in causal epistemology. We are only controlling for observables. We can control for observable features, but we cannot control for things we cannot measure.
            1. Matching: compare units with similar values. Compare subjects with very similar values of covariates.

Still don’t know if we have all the necessary covariates.

Potential for unobserved heterogeneity still exists in matching analyses

What are the reasons people who are so similar get different treatments?

There can be unknowns that don’t exist in data set.

Won’t always show unobserved differences because can’t measure everything. Experimentation allows for unobserved things to be balanced.

* + - * 1. Regression adjustment: multivariate regression. Imposes a functional form on the link between covariates, treatment, and outcome.

Extremely similar to matching. Some underlying move as matching. Compare people with similar values of covariates.

Covariates don’t always have linear relationship between outcome and treatment.

Still don’t know why some subjects got treatment and some did not

Omitted variable bias. Unobserved heterogeneity is a big problem with experiments.

* + - * 1. Propensity scores: model likelihood of receiving treatment. If unit smore likely to get a treatment also have different Y values for other reasons, comparisons between treatment and control will reflect these non-causal differences. Compare units with similar probability of treatment.
    1. Experiments involve interventions. Rather than only observing, as an analyst you get involved in giving treatments.
       1. Randomization
       2. Focus on selection process. How do units get different X values? How did units get into the groups?
          1. Quite often, units have different X values because of pre-existing differences. People and firms make choices for a reason. Typically implausible to believe X is assigned haphazardly, especially if it’s reasonable to think X affects Y.
       3. In experiments, X values are determined by randomization., guaranteeing subjects’ Y values would otherwise be similar if there were no treatment effect. If we are wrong, it can be proven.
       4. Field experiments allow us to infer causal relationships in the real world. Study real-world conditions as closely as possible.

1. Prediction vs. Inference
   1. In previous years, there have been huge advances in predictive accuracy of statistical models. Sometimes you only need to predict Y.
      1. Subtle difference: people most likely to do something won’t necessarily be most likely to respond.
      2. E.g., if person wasn’t shown ad, they would just be as likely to buy, so there isn’t a need to spend on that ad. Just because you can predict, doesn’t mean it’s causal.
   2. Misuses of Predictive Models
      1. Firms often create predictive model scores that predict likelihoods. Predictive models can yield predicted values without clear causal implications.
      2. Predictive models often don’t work out in practice.
      3. Common themes
         1. Treatment effect different from Y
         2. Assumptions can exist without being aware of them
2. Deception and Privacy
   1. Field experiments affected in particular. Intervention is occurring and affecting real people in the real world. Consider ethical implications of choices.
   2. Privacy
      1. Ethical intuitions still evolving
      2. Privacy policies make research difficult
      3. Often want to observe/match data but can’t
         1. Make case for importance of data desired
         2. Find ways to not violate policy such that anonymizing data or randomly assign units in clusters
      4. Think creatively about how to conduct an experiment consistent with a privacy policy if it can’t be changed.
   3. Ethics
      1. Consider costs and benefits of research
      2. Research ethics are cost/benefit analysis
      3. Look at subject’s point of view
         1. Tendency to treat subjects as objects
         2. Consider human impact
      4. Ethical principle: always see your ‘treatment units” as real people.
      5. Common argument: withholding treatment from people in certain situations would be unethical.
         1. Often can’t give treatment to everyone anyway. Consider alternatives. Random assignment and treating everyone possible are not incompatible.
         2. Consider benefits of research. If control group yields good results, it will benefit many more people in the long run.