Observation

- Data analysis allows for making decisions.
- Decisions involve counterfactuals.
    - Existing in the state of the world where one has done X or Y
        - E.g., should women receive hormone replacement therapy (HRT) or not?
        - Should prices be raised or not?
- Observational data: Compare units with different X values.
    - Are women who took HRT better off than those who didn't?

Experimentation

- Experiments involve **interventions**.
- Randomization.
- Focus on selection process.
    - Why do units get different X values?
    - How did units get into the groups?
- Units have different X values because of preexisting 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.

Experimentation (contd)

- X values are determined by randomization, guaranteeing subjects' Y values would otherwise be similar.
- If process is wrong, it can be proven to be wrong.
    - E.g., can be sure that women will have different health outcomes only due to HRT, not other factors
- Field experiments allow us to infer **causal relationships** in the real world.
- Study real-world conditions as closely as possible.

Prediction vs. Causal Inference

- Huge advances in predictive accuracy of statistical models
- Sometimes just need to predict Y
    - "How many shoes will I sell next month?"
        - Decision: How many shoes should I buy? (no experiment needed)
    - "How many website visits will I get?"
        - Decision: Which web hosting to buy? (no experiment needed)
    - "Are men or women more likely to buy my product?"
        - Decision: who to market to? (**Experiment may be needed!**)
- Subtle difference
    - Person most likely to do something won't necessarily be most likely to respond.

Mistaking Prediction for Causal Inference

- Blake et al.: eBay ads on Google searches for "ebay".
    - Specifically branded search
- Seemingly strong evidence that Google search clicks have great return on investment.
    - People who click often buy.
    - Very strong correlation between number of sales and number of clicks.
    - Statisticians didn't want to decrease variable that seemed to predict sales well.
- If ads weren't shown, would people searching "ebay" end up on ebay.com anyway?
    - Randomly assigned some regions to get ads while others didn't.
    - If attribution model is correct, total sales should go down too.
    - Observation alone predicts $1 spent yields $417.3 in revenue.
        - 4,173% ROI

Mistaking Prediction (contd)

- Experiment shows people who clicked ad would have gone to website anyway.
  - −63% ROI.
  - $1 spent yields 37 cents in revenue.
- Experiment showed there wasn't a causal effect.

Predictions and Decisions

- Example: Women more likely to buy product than men.
  - Should advertise more to women?
  - Can predict effect of advertising

Misuses of Predictive Scores

- Firms often create predictive model scores.
  - Predicts likelihoods
- Predictive models can yield predicted values without clear causal implications.

Magazine Example

- Model for subscription cancellations.
  - Percentage chance of cancellation over next few months
- Discount for people likely to cancel.
  - Problem: People may not be responsive to discounts.
- Experiment where random sampling of subscribers received discount.
  - Heterogeneous treatment effect
- Only way to be sure is by running intervention.

Voting Example

- Likelihood of voting for Republican candidate.
  - Idea: Target "moderates" (40–60) with persuasive appeals.
  - 40–60 not moderates, just people we are bad at predicting.
  - Even if person is a moderate, doesn't mean he or she would be receptive to appeals.
- Predictive models often don't work out in practice.

Common Themes

- Treatment effect different from Y.
- Assumptions can exist without being aware of them.

Three Techniques

1. Matching
    - Compare units with similar values.
2. Regression adjustment
    - Multivariate regression
3. Propensity scores

- There is no free lunch.

Matching

- Compare subjects with very similar values of covariates.
    - "Among women of the same race, with similar incomes, blood pressure, height, and weight..."
    - "...those who take HRT are less likely to get cancer than those who don't."
- Still don't know if we have all the necessary covariates.
- Potential for unobserved heterogeneity still exists.
- What are the reasons people who are so similar get different treatments?
- There can be unknowns that don't exist in dataset.

Regression Adjustment

- Extremely similar to matching.
- Imposes a functional form on the link between covariates, treatment, and outcome.
- Example: People who weigh more are more likely to take HRT.
    - Remove effect of weight by adjusting for it.
- Same underlying move as matching.
- Covariates don't always have linear relationship between outcome and treatment.
- Compare people within similar values of covariates.
- Still don't know why some subjects got treatment and some did not.

Omitted Variable Bias

- Unobserved heterogeneity big problem with experiments.
- Incinerator example.
    - Researchers had done regression adjustment (i.e., "controlled for").
- Matching won't always show unobserved differences.
- Can't measure everything.
- Experimentation allows for unobserved things to be balanced.

Propensity Scores

- Key challenge in causal inference is potential connection between **likelihood of treatment assignment** and **outcome**.
  - If units likelier to get treatment also have different Y values for other reasons, comparisons between treatment and control reflect noncausal differences.
- Propensity scores estimate likelihood of receiving treatment directly.
  - Strategy: Compare units with similar probability of treatment.

Example

- Overweight rich women and underweight poor women have a similar chance of receiving treatment, so compare those groups with and without HRT to each other.
  - Model suggests similar likelihood of receiving treatment.
  - If probability of treatment is known, can get unbiased causal effects.
  - Problem remains: We don't know all the reasons why people get treatment.

Example (contd)

- Propensity score can be wrong for many units.
  - E.g., underweight poor women have 80% chance of treatment.
  - Possible some have 99% chance of treatment due to poor health.
  - Other group has 10% chance.
  - Therefore, unclear what is true chance of being treated.
- Another way to do matching.
- Have all reasons for treatment been measured?

Common Themes

- Controlling for observables
- Are there differences between the kind of people who get treatment and don't that we didn't or can't observe?

Tremendous Effort

- Bertrand and Mullainathan:
  - Thousands of fake resumes
  - Thousands of employer listings
- People don't want to do careful research because it's difficult and requires effort.
- Doesn't feel "fancy."
- Right kind of data often hard to get.

Death in the Time of Cholera

	Number of Houses	Deaths From Cholera	Deaths per 10,000 Houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

Snow and Cholera

- Hypothesis: Disease isn't spread through "miasma."
  - Contended that cholera is a waterborne disease
- Ideal experiment: Randomly assign houses to water companies.
- Natural experiment existed.
  - Pipes were laid many years prior in same neighborhoods.
  - Arbitrary who has which water company.
- Knocked on doors to determine people's water company and if they had cholera.
- Nothing "fancy" about table.

Making the Effort

- Put onus on those making assumptions.
  - Why do units get their X values?
  - What determines which units get in groups being compared?
  - Why believe an artificial setting speaks to the setting that's important?
- Some people will say it's impossible to do an experiment that will rigorously answer the questions.
  - Take this as a challenge!
- Think carefully about how to conduct an experiment that will answer big question.
- Worth the work to do careful research.
  - Will say they can't help
  - Will be surprised people are cooperating
  - Will fight findings
- Worth the time and effort.

Deception and Privacy

- Field experiments affected in particular.
  - Intervention is occurring.
  - Affecting real people in the real world.
- Consider ethical implications of choices.

Food Poisoning Letters

- Fake letters sent to restaurants, claiming food poisoning.
- Testing customer-service responses.
- Restaurant employees were fired erroneously.
- Professor conducting study got in big trouble.

Bertrand and Mullainathan

- Measured racial discrimination in job market.
- Sought to quantify effects of race during hiring process.
- Shoe leather research.
- Firms receiving fake resumes were misled and had time wasted.
- People were unknowingly participating in study without giving consent.

Privacy

- Ethical intuitions still evolving.
- Privacy policies make research difficult.
- Often want to observe/match data but can't.
    - Make case for importance of data desired.
    - Find ways around policy.
        - E.g., anonymizing data.
        - Randomly assign units in clusters.
- Think creatively about overcoming privacy policies.

Ethics

- Consider costs and benefits of research.
- Research ethics are cost/benefit analysis.
- Look at subject's point of view.
    - Tendency to treat subjects as objects.
    - Consider human impact.

Ethics (contd)

- Argument: Withholding treatment from people in certain situations would be unethical.
    - E.g., bed nets to protect people from malaria.
    - Often can't give treatment to everyone anyway.
        - Consider alternatives.
        - Random assignment and treating everyone possible are not incompatible.
    - Consider benefits of research.
        - If control group yields good results, it will benefit many more people in the long run.

Experiments Are Changing the World

1. Development
2. Politics
3. Conservation
4. Business

Development Economics

- See books by Karlan and Appel, Banerjee and Duflo.
- How do we increase education?
  - Provide uniforms to girls?
  - Ask teachers to take pictures of themselves?
  - Deworm kids at school?
  - Give cash to families?
- Prior to 2000, most development programs were never really tested.
- With limited resources, allocate randomly.
  - Can know which pilot programs to expand with additional funds
- Without experiments, no way to know the counterfactual.

Politics

- Persuasion and mobilization of voters and volunteers.
- How do we register minorities to vote and turn them out?
- How do we make sure voters hold elected officials accountable for corruption?
- Which governance structures protect minority rights?
- How can activists affect politicians' behavior?
- Questions can start to be answered based on science rather than philosophy.
- Experimentation has transformed political world.

Conservation

- Typical approach: blandishments to conserve.
- Opower sends mail comparing neighbors' power use.
    - Had large effect on people's conservation
    - Frequency of mailings?
    - Amount of social judgment?
    - Effect diminished after several months; new mailings needed
        - Optimal number of mailings to preserve "shock value?"