Importance of Experimentation

- Big data is nice, but majority is not very useful.
  - Does not answer causal questions (e.g., "does x cause y?")
- Having the right kind of data is key.
- Quantity cannot make up for quality of data.
- Experiments can generate the right kind of data.
  - Example: randomized controlled trial.
    - Treatment group: given new cancer drug to determine if it works
    - Control group: given placebo
  - Results from a careful randomized experiment teach us more about causality than do petabytes of observational data.

Example: Hormone Replacement Therapy (HRT)

- Read this 2007 ***New York Times*** article.

Think about the following questions as you read.

- What is the difference in methodology between the two studies?
  - Nurses' Health Study: HRT reduces heart attacks (among postmenopausal women).
  - Women's Health Initiative: HRT increases heart disease (as well as strokes and breast cancer).
- Which one do we believe, and why?

Why Do We Prefer the WHI Results?

- Nurses' Health Study: epidemiological (observational)
  - Big sample sizes.
  - Found statistically significant correlations.
  - "Women who choose to take HRT are less likely to have heart attacks."
- Women's Health Initiative: designed experiment
  - Smaller sample sizes.
  - Clearer conclusions.
  - "Women who we randomly assign to take HRT are more likely to have heart attacks."
- NHS study: Those who choose to take HRT have different health conditions.
- WHI experiment: no preexisting health differences in chosen populations.

The Oat Bran Example

- Potential "invisible" factors affecting results
  - Can be avoided by doing experiment with identical populations

Topic Overview for the Week

- Observation vs. intervention.
- What experiments can tell us.
- Kinds of experiments, in natural and social sciences.
- Example: "Magic on the Internet" auction experiments.
- Assignment: Read Gerber and Green, ***Field Experiments***, Chapter 1.

Reading: Key Points

- Causal questions crucial in a variety of areas (e.g., business, public policy, individual decision making).
- Decisions concerned with counterfactuals.
    - The state of the world if we had done X? If we had done Y?
- Causal inference difficult because we can't observe both states of the world.
- Arguments based on intuition/anecdotes usually stalemates.
    - Extend unemployment benefits in a recession?
        - Might settle with an experiment
    - Keep using Google ads?

Reading: Key Points (contd)

- Causal questions settled with experiments in a way that avoids stalemates.
- Causal questions harder to get correct in social science than in physical science, because of heterogeneity.
    - All electrons the same.
    - No two human minds or bodies the same.
- Should be skeptical of causal inferences based on observational (i.e., nonexperimental) data because of the possibility of unobserved heterogeneity.

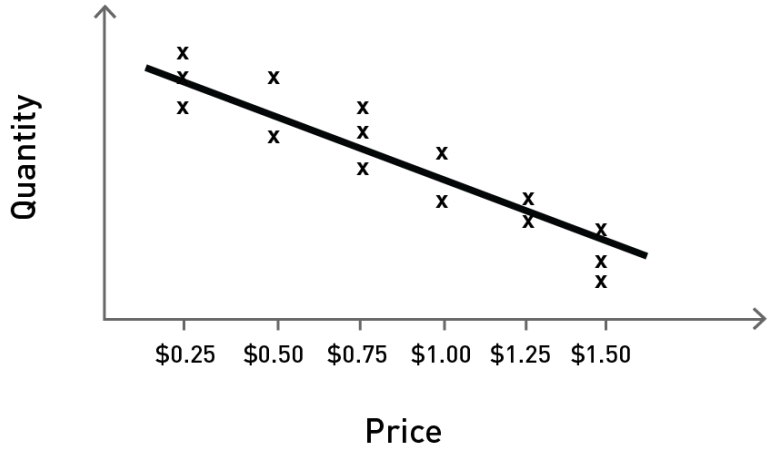
Examples of Causal Questions

- Does boiling drinking water prevent people from contracting cholera?
- Do mandatory seat belt laws reduce traffic fatalities?
- By how much does TV advertising increase purchases?
- Does Chanel No. 5 have a downward-sloping demand curve?
- Does having children give people a more satisfying life?

Definition of Experiment

- **Experiment:** an intervention that creates variation in order to teach us about causal questions

Lemonade Stand Example



- We can't know anything about the demand curve if the price remains the same.
- To learn about the demand curve, we could deliberately vary the price.
    - Variation is crucial for making causal inferences.
- We deliberately create variation when we run experiments.

Intervention and Randomization

- Experiments do not have to involve randomization.
- Intervention is the key element of an experiment.
- Randomization can be difficult to implement.
    - Example: randomly assigning price for each customer
        - Might take too much time
        - Might irritate customers
- We learn more with deliberate variation than without it.

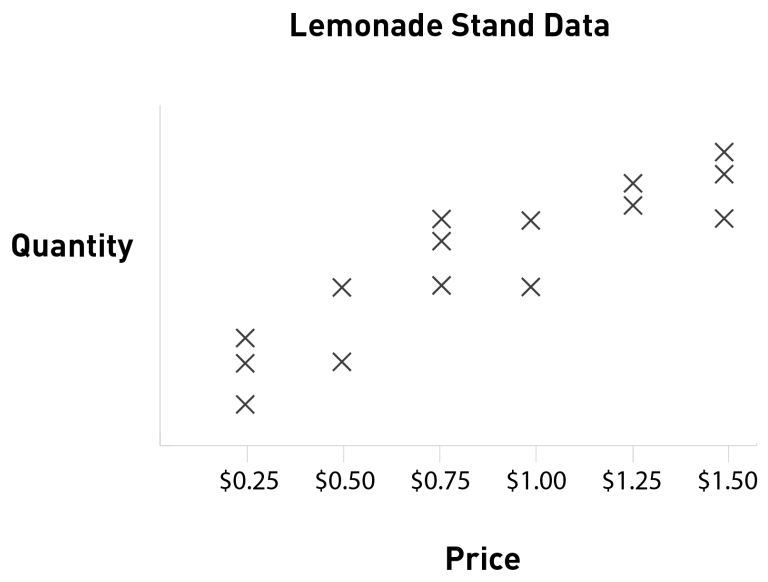
Why Is Randomization Useful?

Consider the lemonade stand example:

- Charging a different price every day
- Charging more on a very hot day, and more lemonade sold
    - Temperature a confounding factor: like price, also affects lemonade sales.
    - Might wrongly assume the price increase (rather than the high temperature) caused the increase in sales.
- Won't learn true effects of price change if any factor that also influences sales is correlated with price change.

Why Is Randomization Useful? (contd)

- We can control for confounds by repeating the experiment many times.
- The treatment—price—should be independent of everything else that might influence lemonade sales (e.g., temperature, sun, holidays).
    - Randomization guarantees this independence.



Pitfalls of Naturally Occurring Variation

- Possible pitfalls in making causal inferences from observational data
  - Nurses' Health Study on hormone replacement therapy (HRT)
    - Possible story: Women receiving hormones generally tended to care more about their health/follow doctors' recommendations.
    - Lower incidence of heart attacks possibly caused by these factors (rather than HRT).
  - Your daughter's unusual lemonade stand data
    - The higher the price, the more lemonade sold.
    - Turns out, higher prices charged when the temperature was higher.
    - Temperature a confounding factor in this case.

What Is a Natural Experiment?

- This is an unpopular term with your textbook authors.
- A natural experiment is naturally occurring data that a researcher argues have the same properties as a true, controlled experiment.

An Analogy: Herschel's Garden

- William Herschel: astronomer who discovered Uranus
- Experiments on stars not possible but one can still learn something by studying various stars at different stages of development.
- **Herschel's garden:**idea of viewing the night sky as a garden that features the same types of "plants" at different stages of development.
    - An idea that can be applied in the social sciences.
    - Natural experiments in the social sciences difficult, however, because of unobserved heterogeneity.

Observational Studies

- Best: data from "naturally occurring experiments" (i.e., situations where variation was produced by something like random assignment)
    - Example: charter school lottery deciding applicant acceptance
        - Data about applicants possibly analyzed to study causal effects of charter school education
        - Same characteristics in both groups, those who attended the charter school and those who did not
    - Example: Vietnam draft lottery that caused some people to get more college education
        - Group that got more college education otherwise identical to group that did not

Observational Studies (contd)

- Less ideal: decent reasons to believe that those who received an intervention are otherwise identical to those who did not
    - Example: snowplows clearing streets on alternate days —effect on business foot traffic?
        - Days of the week not necessarily identical in terms of foot traffic
    - Example: minimum wage increase in New Jersey but not in Pennsylvania—effect on fast-food employment?
        - Minimum wage in New Jersey not raised randomly
- Bad: no good reasons to believe that those who received an intervention are otherwise identical to those who did not
    - Examples: child deliberately charging more for lemonade on hot days; women receiving HRT caring more about their health

Magic on the Internet

- Using field experiments to test equivalence between auction formats.
- Observational data proved unreliable.

Early Online Magic Auctions

- People trading Magic cards via Usenet newsgroup.
- New transaction mechanism (before eBay).
    - Cash used in trades.
    - How to value Magic cards?
- Beginning of Magic auctions.
- Bids updated via e-mail.
- Cards sold if no price increase in three days.

Why Do This Experiment?

- Auction Magic cards in a controlled experiment to learn about auction theory.
- Why is auction theory important?
- Why study revenue equivalence between auction formats?
    - Which format makes more money (e.g., ascending-bid format vs. sealed-bid format)?
        - Should raise same amount on average

Lab Research on Auctions

- Students bid on fictitious good.
- Assign different values to different subjects.
- Each subject knows only her own value, and the probability distribution of others' values
- Can research ascending-bid auction vs. sealed-bid auction.
- Found interesting violations of theory.
- Opportunity to study behavior in preexisting auction market.

Vickrey: Auction Formats

1. English auctions: ascending bids
2. Sealed-bid auctions: one-time highest bids
3. Dutch auctions: descending price clock
4. Vickrey's "second-price" auctions: second-highest bid determines winner's price (like eBay proxy bidding)

Revenue Equivalence: Two Kinds

- Strategic equivalence: strong prediction.
  - Dutch and first-price auctions are strategically equivalent.
    - Same amount of information
  - English and second-price auctions are strategically equivalent under "private values."
    - Dominant-strategy mechanisms for truthful revelation of valuations
      - I.e., regardless of others' strategies, my optimal strategy is the same: bid my maximum willingness to pay.
- General revenue equivalence is weaker.
  - Expected revenue of all four formats should be the same if risk-neutral.

Lab Research: Violations to Theory

- Cox et al. (1982, 83):
  - First-price auctions raise more revenue than Dutch auctions.
- Kagel et al. (1987, 93):
  - With private values, subjects overbid in second-price auctions.
    - Yielded higher revenue than in English auctions
- First/Dutch revenues higher than English/second.
  - Due to risk aversion?

Field Experiments

- As in traditional observational study:
  - Auctions for real goods by people accustomed to bidding for them
- As in lab experiments:
  - Treatment vs. control
    - Two different formats with the same good and bidders
- Cannot control valuations or risk preferences of subjects

Background on *Magic: The Gathering*

- First sold in July 1993, with a first printing of 10 million cards.
- To date, well over 1 billion cards have been printed.
- Estimated 1995 wholesale revenues: over $100 million.
- Cards come in random assortments, which generates a large aftermarket.
- Real-world laboratory.

The Experiment

- Four pairs of auctions designed for within-card comparisons
  - Auctioned the same card twice (e.g., in both Dutch and first-price auctions)
- Sealed-bid auction:
  - One week to submit bids
- English auction:
  - Bid anytime
  - Daily update on each card's high bid
  - Three days with no bid raise to end it

The Experiment (contd)

- Dutch auction:
  - Start at high price
  - Announce decrease of 5% to 10% each day via e-mail
  - Same bid increments as other auction formats
- To control for order effects, each experiment was run twice.
  - FD set: First-price sealed-bid auction followed by Dutch auction.
  - DF set: Reverse of above order.
  - FD and DF sets run at same time.

Result 1: Card-Level Data

- Violates Dutch/first-price revenue equivalence.
- FD and DF experiments: 173 matched pairs of cards.
  - Dutch format: 122 yielded higher revenue.
  - First-price format: 34 yielded higher revenue.
- On average, cards yielded $0.32 more in Dutch auction (24% of value).
- Signed-rank test: price-per-card differences were highly significant.
- No qualitative difference between FD and DF results. That is, order of treatment did not matter.
- Opposite of violation found in lab.

Result 2: Bid-Level Data

- Treatment unit: what constitutes an observation. Here: an individual bid, versus a card receiving multiple bids.
- Bid-level data weakly support violation of strategic equivalence.
  - Theoretically, Dutch and first-price auctions should get same bids.
- Compare bids by same bidder in two matched auctions.
- Data censoring prevents observation of most strategies in Dutch auction.
  - Comparisons not possible for every bidder in both auctions.
- Of 38 observations with bids observed both in Dutch and first-price auctions:
    - 30 favored the Dutch, with mean difference of $2.52.
    - Four favored first-price, with mean difference of −$0.50.

Result 3: English Auction

- Seems to produce slightly more revenue than second-price auction
- Card-level data: 164 pairs observed
  - 1.8% more revenue than second-price auction (not statistically significant)
  - Cannot reject revenue equivalence with this data
- Bid-level data: 112 matched bid pairs
  - 75 cases had higher bids in English.
  - 29 cases had higher bids in second-price.
  - On average, English bid levels were 3% higher.
- Point estimates opposite to lab results
  - Limited statistical power

Result 4:
First/Dutch vs. English/Second

- See how auction revenues deviate from reference price ("cloister price") for each good.
- Pool together Dutch/first-price data and English/second-price: total of 370 observations.
- On average, DF auctions raise at least 12% more revenue than ES auctions.
  - Statistically significant difference (at 90% level)
- Lower difference for higher-priced cards.
  - Difference is zero at about $13.
- Results consistent with earlier lab research.
- Could this be risk aversion?
  - If so, why is the effect smaller for higher-priced cards?

Conclusions

- Revenue ranking:
  - Dutch > first-price > English > second-price
- Field data: opposite violations of FD and ES strategic isomorphisms from lab results
- Same FD > ES effect as lab

Questions

- What caused results to be different from those in the laboratory?
  - Real goods vs. cash payoffs?
  - Simultaneous vs. sequential?
  - Clock speed?
- New laboratory experiment (Katok and Kwasnica, 2003):
  - Slower clock speed leads to higher revenue in Dutch auctions.

Importance of interventions:
holding secondary factors constant and trying alternate methods

The Spirit of Experimentation

- Important to try something new and uncertain.
- Be willing to live with risk of short-term loss.

Questions

- What is an experiment?
- Was this an experiment?
- Would Gerber and Green be skeptical? Why?
- Do you believe the results?
- Are the results representative of other auction environments?

Intervention vs. Observation

- Example: U.S. Forest Service auctioning off timber logging rights.
  - Used both sealed-bid and English ascending-bid auctions.
  - Analysis: One format got higher revenue than another.
  - Results inconclusive due to timber quality factor.
    - Not observed in original data
- Unobserved secondary variables often affect outcomes.
- All secondary variables must be equal to focus on subject.

What to Remember

- Causal questions cannot be answered without intervention (i.e., variation).
- Systematic differences may exist between those who do and do not receive interventions (if naturally occurring).
    - The importance of random assignment
- Difficult when studying people.
    - Make choices about which interventions to receive.
    - Uncontrollable differences among people (i.e., heterogeneity).