Observational studies:

In utero influences

March 14, 2016
First draft deadline is this Friday

- Needs to be at least a mockup of your paper
- With sections that include complete sentences
- At least one table or figure produced and discussed
- Be sure that the 3 sentences from your topic appear in the draft, whether verbatim or updated

1. Question you’re asking
2. Data
3. Answer you expect
Just do it

• We’ve talked about a lot of complex stuff in this class!

• And there will be more complex stuff

• You guys are smart, and you’re also not superhuman

• Your term papers might be works of art, but they don’t have to be

• Don’t worry about designing a perfect study, but try to think about what a perfect study might look like
Today’s agenda

• Review/finish Bootstrap stuff from last time
• Common challenges
• Angus Deaton and RCTs
• Some quick questions about the reading
• More in depth on the reading
• Data exercise
What did you find?

• **Group 1** looked at health of high school grads vs. less than high school (Aaron, Terry, Ravi, (Manish), Morgan)

• **Group 2** looked at health of college grads vs. less than college (Dosbal, Virginia, Emily, Narek)

• **Group 3** looked at numdogs of men vs. women (Shilpa, Hamilton, Sonja, Charles, Shaili)
health of high school grads vs. less than high school

Point estimate of difference: -0.5815865
health of college grads vs. less than college

Point estimate of difference: \(-0.6233746\)
numdogs of men vs. women

Point estimate of difference: 0.0799336
Common challenges

• The data have missing values (nan)

• The data … don’t really look right

• What should I investigate?
Common challenges

• The data have missing values (nan)

  Check Piazza. Thanks to users

• The data … don’t really look right

• What should I investigate?
Common challenges

• The data have missing values (nan)

  Check Piazza. Thanks to users

• The data … don’t really look right

  Sometimes X or Y or both aren’t “cleanly continuous,” maybe heaped, scatterplots aren’t clear, try averages

• What should I investigate?
Common challenges

- The data have missing values (nan)
  
  Check Piazza. Thanks to users.

- The data … *don’t really look right*
  Sometimes X or Y or both aren’t “cleanly continuous,“ maybe heaped, scatterplots aren’t clear, try *averages*.

- What should I investigate?
  Good question, grasshopper.
  Visit office hours?
The 2015 Nobel Prize in economics and RCTs
Nobel in Economics Given to Angus Deaton for Studies of Consumption

By BINYAMIN APPELBAUM  OCT. 12, 2015

Prof. Angus Deaton, a British economist, was awarded the Nobel Memorial Prize in Economic Science on Monday for improving the accuracy of basic economic gauges, including measures of income, poverty and consumption.

Professor Deaton, 69, of Princeton, is best known for his insight that economic averages
Among many things, Deaton has also written about randomized controlled trials

- In the *Journal of Economic Literature*, 2010: RCTs are excellent tools that also have limitations; theory matters

  - Experiments that measure $\partial y/\partial x$ without theoretical analysis sidestep heterogeneity in results & behavior

  - Heterogeneous behavior of humans raises questions whether experiments rigorously separate control from treatment

  - Study findings aren’t automatically relevant to other contexts where circumstances may be different
What RCTs do and don’t do

\[ y = \alpha + \beta x + \varepsilon \]

- RCTs rigorously show the *average* treatment effect (\(\beta\)) of \(x\) on \(y\) between two groups

- But they don’t tell us about differential \(\beta\)’s within the treatment group: heterogeneity

- The FDA will only approve a drug based on results supporting ex ante analytical plans

- Famously, ISIS-2 showed aspirin helped after heart attacks, except if you were a Libra or a Gemini
Strong words for RCTs in development

• Background: Development economists like Esther Duflo (MIT) have pushed RCTs in developing countries

• “[I]f the World Bank had indeed randomized all of its past projects, it is unlikely that the cumulated evidence would contain the key to economic development

• “For an RCT to produce useful knowledge beyond its local context, it must illustrate some general tendency, some effect that is the result of a mechanism that is likely to apply more broadly.”
RCTs without theory are similar to “trial and error” that may not advance scientific knowledge

• Some see experimentation during Enlightenment as central for escaping poverty and disease, but

• “In the end, many problems were simply too hard to be solved without theoretical guidance

• “It took scientific understanding to overcome the heterogeneity of experience which ultimately defeats trial and error

• “We are unlikely to banish poverty in the modern world by trials alone, unless those trials are guided by and contribute to theoretical understanding.”
Question 7.1

What’s clearly bad for baby?

A. Alcohol
B. Caffeine
C. Smoking
D. All of these
E. None of these
Question 7.2

Suppose a study found that heavy drinking was bad for baby. What might you expect to find is true about the moms who report heavy drinking?

A. They’re wealthy moms who can afford it

B. They’re just like the moms who don’t drink heavily

C. They probably also use drugs heavily

D. They’re probably heavier than other moms
Public Service Announcement

• Smokers are people just like the rest of us

• Everyone deserves respect

• The surgeon general of the U.S. warns about the multidimensional risks of smoking

• Smoking is addictive

• We approach the study of smoking with the goal of understanding health disparities, not blaming anyone
Question 7.3

Think about other students in your high school graduating class. Did you know students who were smokers and students who were nonsmokers?

A. Yes, I knew both types

B. No, I only knew one type of student
Question 7.4

Think about the smokers among other students in your high school graduating class. If they were to quit smoking, do you think they would be the same on average as nonsmoker students?

A. Yes, on average they’d be the same if they quit

B. No, they’d still be different

C. I don’t know
Oster, *Expecting Better*, and What To Do When There Are No RCTs

- Wars and other disasters may deprive moms and fetuses of nutrition (fetal programming hypothesis)
- There are no other randomized controlled trials
- Thank goodness!!
- But suppose you’re a pregnant health economist whose doctors tell you all sorts of advice
- Now what?
Observational studies

- A lot of research starts from observation
- But there is smart observation, and not-so-smart
- Lots and lots of characteristics and behaviors are related to one another
- If we want to know a particular $\frac{\partial y}{\partial x}$ but can’t randomize $x$, it’s critical to hold other $z$’s constant

$$y_i = \alpha + \beta x_i + \delta z_i + \varepsilon_i$$
Study design: Following individuals over time

- Suppose we’re interested in how a bad $x$ affects an outcome $y$

- We could observe people over time, measure $x$ & $y$, then compare $y$ across high-$x$ and low-$x$ people

- How useful will this be? Depends on
  - How homogeneous the sample of people was
  - Whether $x$ is correlated with something else that we’re not measuring (more likely with less homogeneity)
Alcohol & pregnancy

• One of my favorite passages in the chapter is on page 51:
  - One phrase I kept coming across was "no amount of alcohol has been proven safe."
  - [T]oo much of many foods can be bad. If you have too many bananas (and I mean a LOT of bananas), the excess potassium can be a real problem. But no doctor is going around saying "No amount of bananas have been proven safe!" He'd be laughed out of the medical conference.
  - [E]vidence … leads us to conclude that binge drinking is problematic. But if you are willing to conclude that, why wouldn't you be willing to conclude that light drinking is fine? That is what the evidence shows.
“Evidence” from studies of alcohol and pregnancy: The sample really matters

• **External validity** is how a study’s results may apply to other circumstances

• There are no RCTs that vary drinking across moms

• But there are studies of moms drinking
  - In the U.S., where it’s become a huge taboo
  - In other countries, like Australia, where moderate drinking is more quotidian
Drinking in Pregnancy and Behavior Problems Among 2-Year-Olds

- Level of behavior problems with no drinking
- Level of behavior problems with occasional drinking (≤1 drink per week)
- Level of behavior problems with light drinking (2-6 drinks per week)
- Level of behavior problems with moderate drinking (7-10 drinks per week)

Versus

Percentage of children with behavioral problems
They measured IQ with a test called Raven's matrix. It works like most IQ tests in that higher scores are better, and the test is designed so that the average person will score 100. Here's the data:

<table>
<thead>
<tr>
<th>Drinking Categories</th>
<th>Raven's Matrix Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>No drinking</td>
<td>103</td>
</tr>
<tr>
<td>Less than 1/2 glass per day</td>
<td>104</td>
</tr>
<tr>
<td>1/2-1 glass per day</td>
<td>105</td>
</tr>
<tr>
<td>More than 1 glass per day</td>
<td>106</td>
</tr>
</tbody>
</table>

Just as in the study of behavior, there is no evidence here to suggest that the children of light drinkers are worse off than those of women who drink nothing. In fact, their scores are higher on average (although these results are not statistically significant—they may just reflect random variation).

The researchers concluded there is no evidence of worse test performance, even among the children of moms who have a drink or more per day.

Raven IQ score (more is better)
This is not to say that one cannot unearth studies that find that light drinking is a problem.

The issue is that these studies are very deeply flawed.

One of the very nice things about the previous studies ... was that ... women who drank different amounts were not that different in other ways.
Alcohol, context, and omitted variables

• The cited study by Sood et al. (Pediatrics 2001)
  - Data from Fetal Alcohol Research Center, Wayne State U
  - 90% of moms were African American
  - Of moms who reported no alcohol use, 18% reported cocaine use. Of moms who drank, 45% reported cocaine
  - “These data suggest that no alcohol during pregnancy remains the best medical advice.”

• Oster: “At this point, I threw that paper in the trash. Maybe if I was wondering about combining my end-of-day glass of wine with cocaine it would be useful.”
Papers that omit variables still get published

• A literature shows reduced child neurocognitive outcomes associated with advanced paternal age

• Sperm can mutate with age, but no RCT evidence exists

• What else is paternal age correlated with?
  - Maternal age (controlled)
  - Birth order and family size (not controlled), Mother’s education (not controlled)

• Edwards and Roff (2010) show many “effects” are bogus
Caffeine & pregnancy

• Turns out there was an RCT in Denmark!

• Also RCT using lab rats:
  - Miscarriage in rats with 250mg per kg per day
  - For humans, translates to 60 cups of coffee per day

• A lot of observational studies of self-reported caffeine use and pregnancy outcomes
Nausea is good?


Women reporting vomiting were

- less likely to experience miscarriage or stillbirth ($P = .002$)
- and delivery before 37 weeks' gestation ($P = .004$)
- but there was no difference in infant birth weight between mothers with and without vomiting ($P = .48$).”
Share of Women Reporting Vomiting, by Pregnancy Week

Nausea, coffee, miscarriage, and omitted variable bias

• What moms drink depends on how they feel

• Nauseous moms won’t drink coffee

• But nausea isn’t a binary all-or-nothing kind of thing, although some studies measure it that way

• If we can’t adequately control for nausea, it’s plausible that all the “effects of coffee” on miscarriage are actually the lack of nausea
Clever analysis of coffee drinking results (p. 58)

• It’s logical to assume that the amount of caffeine or coffee should matter for outcomes

• Then we’d expect reductions in the level to matter, but also the starting and ending levels

• If coffee drinking instead is a proxy for no nausea, then a reduction means nausea happens

• Studies that find only the reduction in coffee is important, not the levels, are probably showing us the effects of nausea, not caffeine
Clever study: Decaf

• Why would coffee cause miscarriage? Caffeine

• What else is associated with miscarriage? Lack of nausea

• Moms who aren’t nauseous will probably also drink more coffee because they’re not nauseated

• Oster (p. 59): Let’s examine moms who drink decaf coffee

• Turns out that’s as strongly associated with miscarriage!

• It’s not the caffeine, apparently. Lack of nausea? Maybe
Danish RCT on caffeine! (Bech et al., BMJ 2007)

- Recruited of 1,207 pregnant women who reported 3+ cups of coffee per day
- Researchers asked them to replace with instant coffee, randomized between decaf and regular
- The study found no statistically significant differences in outcomes between control & treatment
- “Conclusion. A moderate reduction in caffeine intake in the second half of pregnancy has no effect on birth weight or length of gestation.”
Smoking & pregnancy

- Oster “gets us” by starting off with a bunch of cross-sectional associations between smoking and bad outcomes.

- She knows that we’ll say, “what about omitted variables like the mother’s socioeconomic status?”

- Then she lowers the boom:
  - Because smoking is (now) known to be bad,
  - RCTs in which smoking moms are randomly chosen for smoking cessation programs exist.

- (Only, ask yourself: Is a smoking mom in a program who then quits really the same as a never-smoking mom?)
Smoking is bad, m’kay

• Not only is it bad to produce carbon monoxide and absorb nicotine

• It’s also really hard to quit

• Oster cites treatment groups that drop from 90% smoking to 80% smoking

• Average treatment effects are thus small because almost nobody successfully quits

• Scaled up, you’d get +14 oz of birth weight for actually quitting!
Smoking and weight (loss)

- http://demog.berkeley.edu/~redwards/LS88/c07_smokeweight.csv

- Let's look at two things:
  - Compare the weight in kg (r8weight) of smokers (r8smoken == 1) and nonsmokers (r8smoken == 0)
  - Examine the weights in waves 8 and 9 in kg of quitters (r8smoken == 1 and r9smoken == 0)