Advanced Quantitative Methods

Experiments

Instructor: Gregory Eady
Office: 18.2.10
Office hours: Fridays 13-15

A few comments for Assignment 1

- Remember to create variables *within* the existing dataset.
  - Don't create multiple datasets, except for graphing
- Try to do all of your cleaning in one place
  - In a typical project, you'd have a specific .R file *just* for cleaning data
- For bar graphs, use percentages (not counts)
- "Percentage point" and "percent"
- "Probability" not "likelihood"

Today

- Interaction terms in OLS
- Experiments
  - What do we mean by causal inference?
  - Why do experimental designs permit causal inference?
  - Analyzing a basic experiment in R

Interaction terms in regression models

Linear probability model predicting whether someone votes for Donald Trump (i.e. basic OLS)

$$vote\_trump_i = \alpha + \beta_1 ideology_i + \beta_2 male_i + \beta_3 ideology_i \times male_i + \epsilon_i$$

Model results

Table 1: OLS regression results of vote for Donald Trump

	Vote for Trump		
	(1)	(2)	(3)
Ideological self-placement	0.183***	0.182***	0.175***
	(0.001)	(0.001)	(0.001)
Ideological self-placement × Male			0.015***
			(0.002)
Male		0.044***	−0.017*
		(0.004)	(0.009)
Intercept	−0.285***	−0.301***	−0.274***
	(0.004)	(0.005)	(0.006)
Observations	39,808	39,808	39,808
R^2	0.459	0.461	0.462
Note:	*p<0.1; **p<0.05; ***p<0.01		

$$vote\_trump_i = \alpha + \beta_1 ideology_i + \beta_2 male_i + \beta_3 ideology_i \times male_i + \epsilon_i$$

- $\beta_1 ideology_i$
  - β_1 is the relationship between ideology and voting for Donald Trump *when* $male_i = 0$
  - Do *not* interpret this as the unconditional "effect" of ideology
  - Statistical significance of this term is only for this conditional case

$$vote\_trump_i = \alpha + \beta_1 ideology_i + \beta_2 male_i + \beta_3 ideology_i \times male_i + \epsilon_i$$

- $\beta_2 male_i$
  - β_2 is the relationship between being male and voting for Donald Trump *when* $ideology_i = 0$
  - Do *not* interpret this as the unconditional "effect" of being male
  - Statistical significance of this term is only for this conditional case

$$vote\_trump_i = \alpha + \beta_1 ideology_i + \beta_2 male_i + \beta_3 ideology_i \times male_i + \epsilon_i$$

- $\beta_3 ideology_i \times male_i$
  - β_3 is the *difference* in the "effect" of ideology for men relative to women
  - i.e. Allows us to ask: When $male_i = 1$ does ideology have a stronger or weaker "effect"?
  - Statistical significance tells you whether the "effect" is indeed bigger or small for men versus women

Nobel in economics goes to David Card, Joshua Angrist and Guido Imbens.

All three winners are based in the United States. Mr. Card, who was born in Canada, works at the University of California, Berkeley. Mr. Angrist, born in the United States, is at M.I.T. and Mr. Imbens, born in the Netherlands, is at Stanford University.

“Uncovering causal relationships is a major challenge,” said Peter Fredriksson, chairman of the prize committee. “Sometimes, nature, or policy changes, provide situations that resemble randomized experiments. This year’s laureates have shown that such natural experiments help answer important questions for society.”

scott cunningham and Arindrajit Dube liked

Peter Hull
@instrumenthull

big day for OLS

8:04 AM · Oct 11, 2021 · Twitter for Android

12 Retweets **196** Likes

But field experiments are nevertheless typically the gold standard

- Blattman & Dercon (2018)
- Enos (2014)
- Bertrand & Mullainathan (2004)
- Munger (2017)

"The Impacts of Industrial and Entrepreneurial Work on Income and Health" (Blattman & Dercon 2018)

- Does industrial work help people in developing countries?
- Randomize who gets a job in a sweatshop
- $\rightarrow$ Adverse effects on health and no effect on income (because informal alternatives)

"Causal effect of intergroup contact on exclusionary attitudes" (Enos 2014)

- Does physical proximity to an out-group affect attitudes?
- Randomizes the presence of Spanish-speaking hispanics at train stations in Boston's predominantly white suburbs
- → Increase in exclusionary attitudes

"Are Emily and Greg More Employable Than Lakisha and Jamal?" (Bertrand & Mullainathan 2004)

- Does race and gender affect labor market outcomes?
- Use what is called an "audit" or "correspondence" experiment
- Randomizes "white"- and "black"-sounding names on CVs
- $\rightarrow$ White names receive 50 percent more callbacks for interviews

"Tweetment Effects on the Tweeted" (Munger 2017)

- Does social sanctioning reduce incivility on social media?
- Randomize a message sent to users who use the n-word on Twitter
- → Decrease in subsequent use of racist language

Tweetment Effects on the Tweeted (Munger 2017)

Fig. 3 Treatments. **a** The treatment—black bot. **b** The bot applying the treatment—white bot

Survey experiments

"The Domestic Political Costs of Soliciting Foreign Electoral Intervention" (Tomz & Weeks, Forthcoming)

Table 1: Dimensions of Treatment Scenarios

Dimension	Randomized Values
Democrat	Joe Biden, Pete Buttigieg, Kamala Harris, or Bernie Sanders
Republican	Mike Pence, Ted Cruz, Ron DeSantis, Mike Pence, or Donald Trump
Country	An *ally* (Australia, Canada, France, Japan, or the UK) or a *nonally* (China, Iran, North Korea, Russia, or Syria)
Request	Endorsement, turnout ads, positive ads, negative ads, or investigation
Quid pro quo	None, diplomatic, economic, institutional, or military

Figure 1: Sample Treatment Scenario

Suppose that ...

- The main candidates in the 2024 U.S. presidential election were Joe Biden (Democrat) and Mike Pence (Republican).
- During the campaign, Biden sent a private message to the leader of China.
- In the message, Biden asked China to buy anonymous online ads criticizing Pence's character and policies.
- Biden promised that, if elected, he would return the favor by signing a trade agreement that would help China's economy.
- China has not yet responded to the message.

If events happened just as we described, then all things considered, which candidate would you prefer?

Conjoint experiments

"How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers" (Bansak et al. 2016)

Two different asylum seekers

Different attributes of the asylum seekers

ATTRIBUTE	APPLICANT 1	APPLICANT 2
Age	21 Years	62 Years
Language Skills	Speaks broken English	Speaks fluent English
Previous Occupation	Unemployed	Teacher
Religion	Christian	Muslim
Consistency of Asylum Testimony	Minor inconsistencies	Major inconsistencies
Vulnerability	Post-traumatic stress disorder (PTSD)	No surviving family members
Origin	Iraq	Pakistan
Reason for Migrating	Seeking better economic opportunities	Persecution for ethnicity
Gender	Male	Male

"How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers" (Bansak et al. 2016)

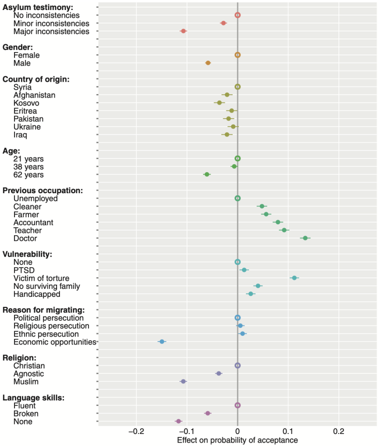

Fig. 2. Effects of asylum-seeker attributes on the probability that respondents accept the asylum seeker. Dots with horizontal lines indicate point estimates with cluster-robust 95% confidence intervals (CI) from linear (weighted) least squares regression. The unfilled dots on the zero line denote the reference category for each asylum-seeker attribute. Table S9 (model 1) displays the underlying regression results.

What if...

- You were not offered a job in a sweatshop?
- You did not hear people speaking Spanish on your metro ride to work?
- Your name did not sound ethnically African-American?
- You were not called out for being racist?

Causal effects

- The causal effect is the comparison between what one might have believed or done had those things not happened
- A "counterfactual":
  - Imagining what an outcome would be in hypothetical world in which something did/didn't happen
- By definition, we cannot observe this counterfactual world
  - i.e. we don't get to see what happens if we assign someone to the control group *and* what happens if we assign them to the treatment group—we only observe one of these outcomes
- This is the fundamental problem of causal inference

Social pressure experiment (Gerber et al. 2007)

Neighbors mailing

30423-3

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C050
THE JACKSON FAMILY
9999 MAPLE DR
FLINT MI 48507

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	
9995 JENNIFER KAY SMITH		Voted	
9997 RICHARD B JACKSON		Voted	
9999 KATHY MARIE JACKSON		Voted	
9999 BRIAN JOSEPH JACKSON		Voted	
9991 JENNIFER KAY THOMPSON		Voted	
9991 BOB R THOMPSON		Voted	
9993 BILL S SMITH			
9989 WILLIAM LUKE CASPER		Voted	
9989 JENNIFER SUE CASPER		Voted	
9987 MARIA S JOHNSON	Voted	Voted	
9987 TOM JACK JOHNSON	Voted	Voted	
9987 RICHARD TOM JOHNSON		Voted	
9985 ROSEMARY S SUE		Voted	
9985 KATHRYN L SUE		Voted	
9985 HOWARD BEN SUE		Voted	
9983 NATHAN CHAD BERG		Voted	
9983 CARRIE ANN BERG		Voted	
9981 EARL JOEL SMITH			
9979 DEBORAH KAY WAYNE		Voted	
9979 JOEL R WAYNE		Voted	

Social pressure experiment (Gerber et al. 2007)

TABLE 3. OLS Regression Estimates of the Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Model Specifications		
	(a)	(b)	(c)
Civic Duty Treatment (Robust cluster standard errors)	.018* (.003)	.018* (.003)	.018* (.003)
Hawthorne Treatment (Robust cluster standard errors)	.026* (.003)	.026* (.003)	.025* (.003)
Self-Treatment (Robust cluster standard errors)	.049* (.003)	.049* (.003)	.048* (.003)
Neighbors Treatment (Robust cluster standard errors)	.081* (.003)	.082* (.003)	.081* (.003)
N of individuals	344,084	344,084	344,084
Covariates**	No	No	Yes
Block-level fixed effects	No	Yes	Yes

Note: Blocks refer to clusters of neighboring voters within which random assignment occurred. Robust cluster standard errors account for the clustering of individuals within household, which was the unit of random assignment.

* $p < .001$.

** Covariates are dummy variables for voting in general elections in November 2002 and 2000, primary elections in August 2004, 2002, and 2000.

The basic framework for understanding experimental results

- Take the average of an outcome between two groups (e.g. proportion who voted)
  - Treatment group receives a stimulus ("Your neighbors will know if you voted")
  - Control group does not receive that stimulus
- However, without randomized assignment:
  - Comparison is polluted by selection bias
  - People often "select into" treatment (e.g. healthy people often the ones who use wellness programs)

Randomization *breaks* selection bias because no one is more/less likely to be assigned to the treatment or control

Thinking in terms of "Potential outcomes"

% budget on water sanitation
if we could observe the counterfactual

Village i	Budget Share $Y_i(\text{Leader} = \text{Male})$	Budget Share $Y_i(\text{Leader} = \text{Female})$	Treatment effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	15	20	5

Potential outcomes

% budget on water sanitation
as we actually observe it in reality

Village i	Budget Share $Y_i(\text{Leader} = \text{Male})$	Budget Share $Y_i(\text{Leader} = \text{Female})$	Treatment effect
Village 1	?	15	?
Village 2	15	?	?
Village 3	20	?	?
Village 4	20	?	?
Village 5	10	?	?
Village 6	15	?	?
Village 7	?	30	?
Average	16	22.5	6.5

Some basic notation

- E[·] Statistical expectation. Think of it just as an average.
- $Y_i(1)$ Potential outcome if received treatment, $Y_i(Leader{=}Female)$
- $Y_i(0)$ Potential outcome if received control, $Y_i(Leader{=}Male)$
- $E[Y_i(1)|T_i = 1]$
  - Expected value of $Y_i(1)$ among actually treated units
- $E[Y_i(0)|T_i = 1]$
  - Expected value of $Y_i(0)$ among actually treated units
- $E[Y_i(0)|T_i = 0]$
  - Expected value of $Y_i(0)$ among actual control units
- $E[Y_i(1)|T_i = 0]$
  - Expected value of $Y_i(1)$ among actual control units

The average treatment effect (ATE)

$$\underbrace{E[Y_i(1)|T_i = 1]}_{\substack{\text{Average treatment outcome} \\ \text{among the treatment group}}} \quad - \quad \underbrace{E[Y_i(0)|T_i = 0]}_{\substack{\text{Average control outcome} \\ \text{among the control group}}}$$

This is just the difference in the average outcome among the treatment group and the average outcome among the control group

Average treatment effect on the treated (ATT)

$$\underbrace{E[Y_i(1)|T_i = 1]}_{\substack{\text{Average treatment outcome} \\ \text{among the treatment group}}} - \underbrace{E[Y_i(0)|T_i = 1]}_{\substack{\text{Average control outcome} \\ \text{among the treated group}}}$$

This is the average treatment effect *among the treatment group*. Treatment group is often a different population.

As we will see later on with quasi-experiment, the ATT is often the best we can get

If no selection bias:

- In a counterfactual world in which no one is treated, we should expect no difference in outcomes between the treatment and control group
- If we do see a difference, then we have selection bias
- i.e., if $E[Y_i(0)|T_i = 1] \neq E[Y_i(0)|T_i = 0]$
- If so, then a difference in means will give us the ATE + the selection bias:
  - Difference in means $=$ ATE $+ E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 0]$

Our goal, then, is to keep everything else equal between the treatment and control groups

- Remove the reason why some receive the treatment and others do not—as we often have in observational data
- The challenge in causal research is typically to find a way to eliminate selection bias
- Model-based inference does this primarily with regression and controls
- Design-based inference does this through explicit randomization or clever quasi-experiments

The law of large numbers

- Draw a sample from a population *at random*
- The average in the *sample* will be similar to the average in the *population*
- The larger the sample, the greater the similarity (less variation).
- The same applies to treatment and control groups.
    - We have a group of people we would like to experiment on
    - Assign everyone to a treatment or control group at random
    - The average on all observed and *unobserved* variables in the two groups will be equal in expectation

The result of randomization

- Individuals assigned to the control and treatment groups are not the same (fundamental problem of causal inference).
- But they are the same on average (in expectation)
- This allow us to easily estimate the average treatment effect
- There is variability, of course, but we can capture that with basic statistical tools (our standard errors)

A quick demonstration of this in R...

We can estimate the treatment effect with just OLS

$$y_i = \alpha + \beta T_i + \epsilon_i$$

- OLS with a binary variable T_i just gives us a difference in means

$$y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i$$

- Can include controls, X_i, to get more precision, because can be imbalances even with randomization
  - But we must declare that we will do this before we run an experiment (in a "pre-registration plan" of the experiment)

Randomization allows for unbiased estimation of causal effects. But this can also break down:

- Exclusion restriction
  - That potential outcomes only respond through treatment assignment (and no other channel)
  - e.g. This would be broken if an aid organization knows about the village experiment and thus intervenes to provide more help to male-led organizations
- Stable Unit Value Treatment Assignment (non-interference)
  - e.g. The treatment of one village with a woman leader does not affect the outcome of another village

Other things to think about:

- Experimenter "demand effects"
  - Research subjects intuiting what the goal of the experiment is and thus responding in a way favorable to the experimenter (but Mummolo & Petersen 2019)
- Researcher / affiliation characteristics
  - e.g. survey experimenter is a man or woman
- Hawthorne effects
  - Knowing one is in an experiment can affect behavior
- Non-compliance
  - Not everyone follows through with the treatment (e.g. wellness programs)
  - But can get a Complier Average Causal Effect
- Clustered treatment assignment (e.g. classrooms)
- Effect heterogeneity (i.e. with interaction term)

Open science and power analysis

- **Pre-registration**
  - p-hacking and the garden of forking paths
  - The file drawer
  - All high-quality experiments are now "pre-registered"
- **Power analysis**
  - Ensures that you don't run a costly experiment with too few subjects
  - "If my best guess is that the effect of my experiment will be of magnitude x, how big of a sample size will I need to find a statistically significant result at least 80% of the time?"
- If you use experiments in your thesis, pre-register your design!
- It is what practicing researchers do, and will look good to an external examiner (as it should)

R Exercise

- Download exercise from the course website

Exercise solutions

```
# What proportion of women are in the treatment group?
mean(D$woman[D$treatment == 1])

# What proportion of women in the control?
mean(D$woman[D$treatment == 0])

# Or
D %>%
group_by(treatment) %>%
summarize(woman = mean(woman))


# What if you wanted to have a control group and two treatment groups, each of
# which is 1/3 of the sample? Where the control is coded "Control", the first
# treatment is coded "Treatment 1" and the third treatment is coded "Treatment 2"
# Use the function "round()" to get whole numbers
D$treatment <- "Control"
D$treatment[(N/3+1):(N*2/3)] <- "Treatment 1"
D$treatment[(N*2/3+1):N] <- "Treatment 2"
```

Exercise solutions

```
# Load in their data. i.e. "Gerber_et_al_2008.csv"
G <- read_csv("Your_File_Path/Gerber_et_al_2008.csv")

# If we want to calculate the effect of the treatments relative to the control
# we need to make it a "factor" variable with factor() and order the
# categories such that "Control" is the first of the 4. The other of each
# treatment doesn't matter. You can make the order something like this:
# c("Control", "Civic Duty", "Hawthorne", "Neighbors")
G <- G %>%
    mutate(messages = factor(messages,
                             levels = c("Control", "Civic Duty",
                                        "Hawthorne", "Neighbors")))
```

Exercise solutions

```
# To analyze the experiment run a standard OLS regression model, where the
# outcome variable is "primary2006" and the treatment variable is "messages".
ols_model <- lm(primary2006 ~ messages, data = G)

# Output the regression results to stargazer
stargazer(ols_model, type = "text")

# Now analyze the experiment running the standard OLS regression model, but
# including the variable "sex" as a control
ols_model_with_control <- lm(primary2006 ~ messages + sex, data = G)

# Look at the coefficients of the model with summary()
# Is there a meaningful difference between the estimated treatment effects
# when sex is controlled for? If not, why not?
summary(ols_model_with_control)
```