

Lecture 1 - Introduction

ECON 4651: Principles of Econometrics for Business and Analytics

Jason Cook
Fall 2020

Instead of waiting in awkward silence for everyone to join, enjoy this live panda footage



Prologue

Who am I?

Jason Cook

- Applied microeconomist and econometrician
- I study the economics of education and public economics.
 - Charter schools
 - Racial segregation
 - Food Assistance: SNAP/WIC



Where can you find me?

- **Office hours:** T/Th 3:30-4:30pm
or by appointment. Schedule
with **Calendly**
- **Email: Don't email (use Slack
Instead!)**

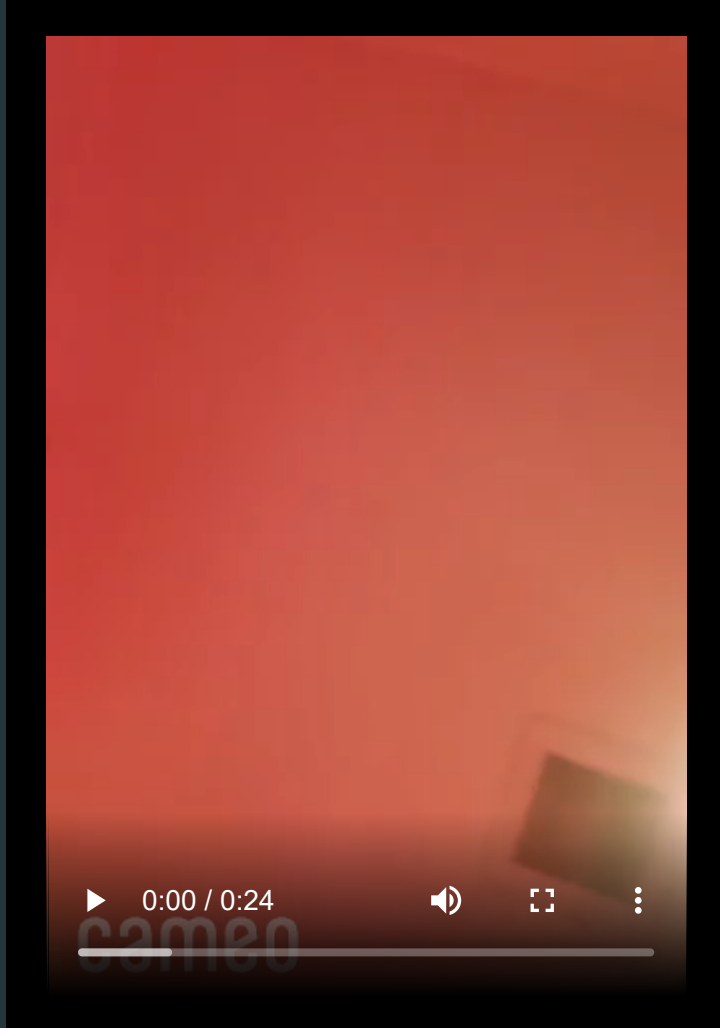


Teaching Assistant

- **Blake Peterman**
- **Email:** u0965622@utah.edu
- **Office Hours:** T/F 4pm-5pm
 - **Zoom ID:** 988 6826 2146
 - **Password:** 4651

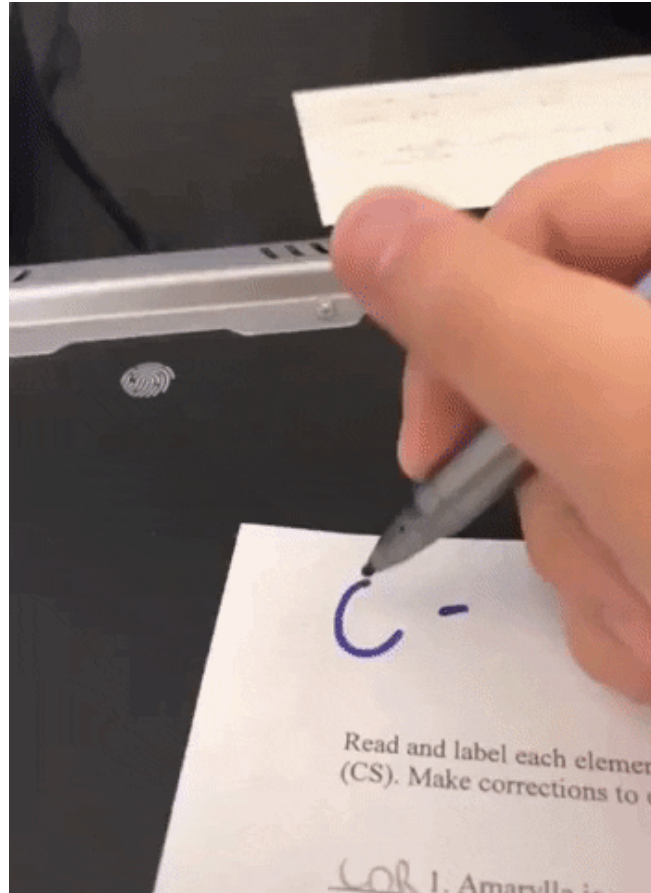


Syllabus



Syllabus

Grading



Syllabus

Grading

- Attendance **(1 pt per class up to 25 pts - 5%)**
 - Watch entire lecture, either by Zoom or Canvas, or combo
 - Lab days, submit Stata do-file for credit
 - Must be completed before Sunday at midnight following the lecture
- 6 Problem Sets **(25 pts each - 25%)**
 - Mix of theory and empirical tasks
 - Optional groups of up to 4 **(graded as a group)**
 - Will drop the lowest-scored problem set

Syllabus

Grading continued...

- 4 Group Assignments **(25 pts each - 20%)**
 - Assigned a group of 5[†]
 - Open-ended empirical assignments
 - Submit report and video presentation
 - Give feedback on 3 other presentations
- Midterm - ProctorU **(100 pts - 20%)**
- Comprehensive Final - ProctorU **(150 pts - 30%)**
- Scores curved within the class
 - This is a challenging course and so curve will most likely be used to move student scores **upward** to achieve a reasonable class distribution

[†]: Contact TA if group member isn't helping and we will contact entire group and may dock points to that member

Syllabus

Contesting Grades

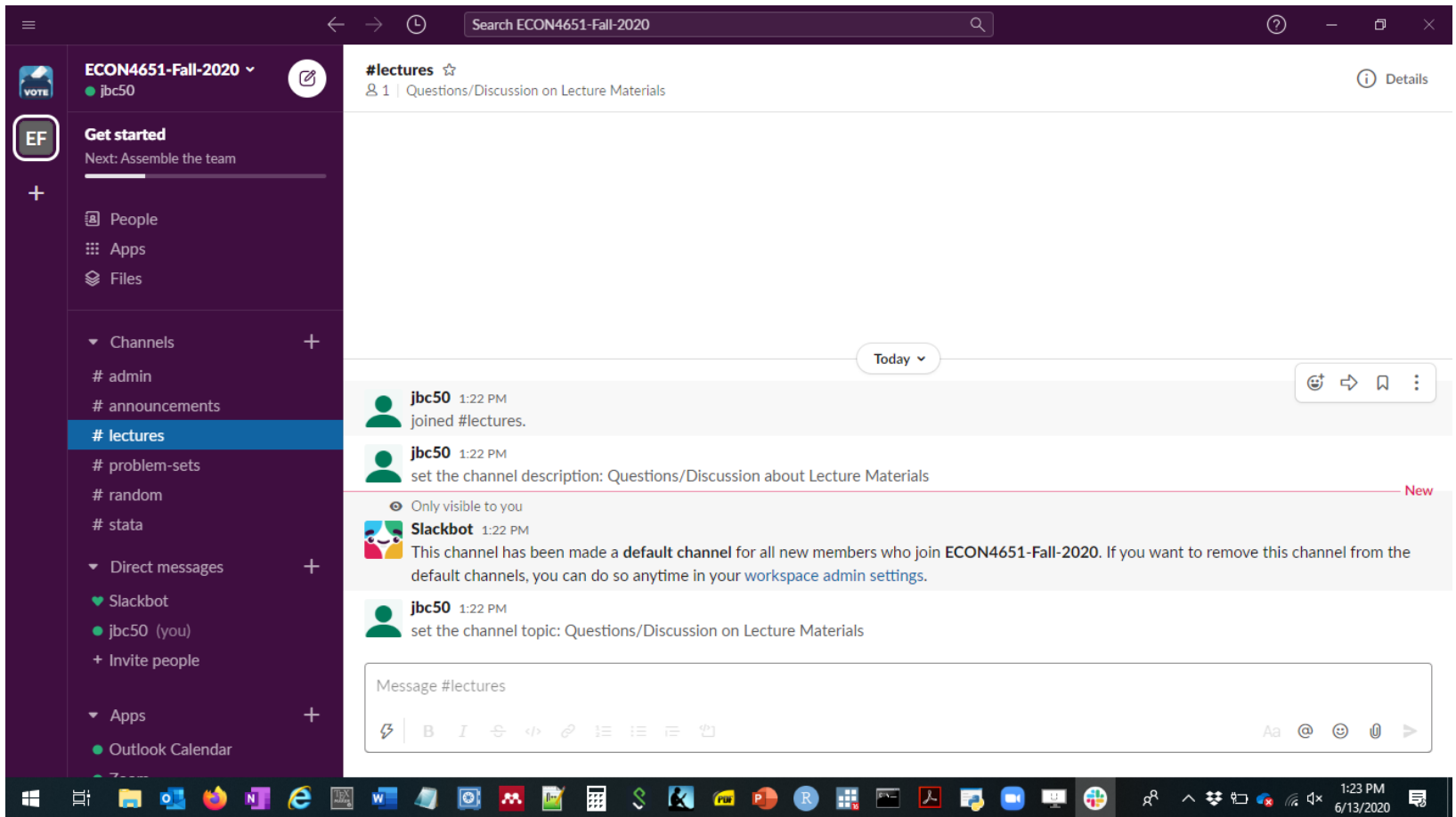
- You have **two weeks** from when returned to contest grades on any assignment
- First, consult with TA to try to come to resolution
- If that fails, then I will regrade
- You must accept my scoring as final, even if it is lower than the original grade
 - In the past, I've been a tougher grader than my TA so factor that in

Slack

Slack

- I will utilize Slack for this course rather than email **(response guaranteed)**
- Please do not email me **(you may not get a response)**
- Create a Slack Account and join our Workspace **Econ 4651: Fall-2020**
- **Getting Started with Slack**
- Be sure to add **profile pictures** as well. It helps to feel connected in these remote semesters.

Slack



- Channels (#) along the left panel organize topics for discussion
- Send Direct Messages in lower left panel

Slack


In Class

- We will periodically use Slack in class for impromptu polling and group activities, so have it ready
 - These will appear in #in-class channel
- You can also post questions about lecture material in #lecture
 - These will be answered by myself, the TA, and your fellow peers (not necessarily during class time)
 - For questions you want an immediate answer to, use Zoom features (described shortly)

Let's practice...

Open up Slack, see you there

Ettiquete

- Respond to comments and questions in **threads**
 - Hover over message and click thread  icon
- Be kind and professional
- Slack is **judgment-free zone**, no such thing as "stupid questions"
- TA and I will monitor Slack to answer questions as well
- To ask a specific question to me or Blake, be sure to tag us in the post using @
- Please do answer other students questions and contribute to discussion
- In fact...

Slack

Bonus

To help facilitate activity on Slack. At the end of the semester, I will subjectively award up to **25 points** (equivalent to one full problem-set) to each student based on their activity on Slack

Activity includes both asking and answering questions, though I'll put more weight on answering



Channels:

- #admin
- #announcements
- #lectures
- #problem-sets
- #in-class
- #random
- #stata

Direct Messages:

- Send direct messages jointly to myself **and the TA** for any personal questions or concerns
 - TA will usually respond, but I'll step in as needed

Zoom

Zoom

Ettiquette

- keep video on (if possible)
- raise hand
- use chat feature
- mic off unless called on

Group Questions

- will regularly split into small groups to discuss content
- Let's practice, just introduce yourselves to your other group members for a few minutes and I'll bounce around



Poll Everywhere

Poll Everywhere

- I'll use poll everywhere for most in-class polling
- These polls are to gauge understanding and will not be for points
- No need to enroll in Poll Everywhere, just click links in Slack channel
 - But enrolling does make it a bit faster for you
- Let's try it out

Poll Everywhere

Motivation

Why study econometrics?

1. Develop **skills that employers value**.
2. Cultivate **healthy skepticism**.
3. Learn about the world using **data**.

Motivation

Why study econometrics?

Provide answers to important questions

- Do minimum wage policies **reduce poverty**?
- Does the death penalty **deter violent crime**?
- Does recreational marijuana **cross state lines**?
- Are recessions **good for your health**?
- How will global warming **affect the economy**?
- Will Donald Trump **win again**?
- Do mandatory mask policies **reduce the spread of COVID-19**?

Econometrics

Most econometric inquiry concerns one of two distinct goals:

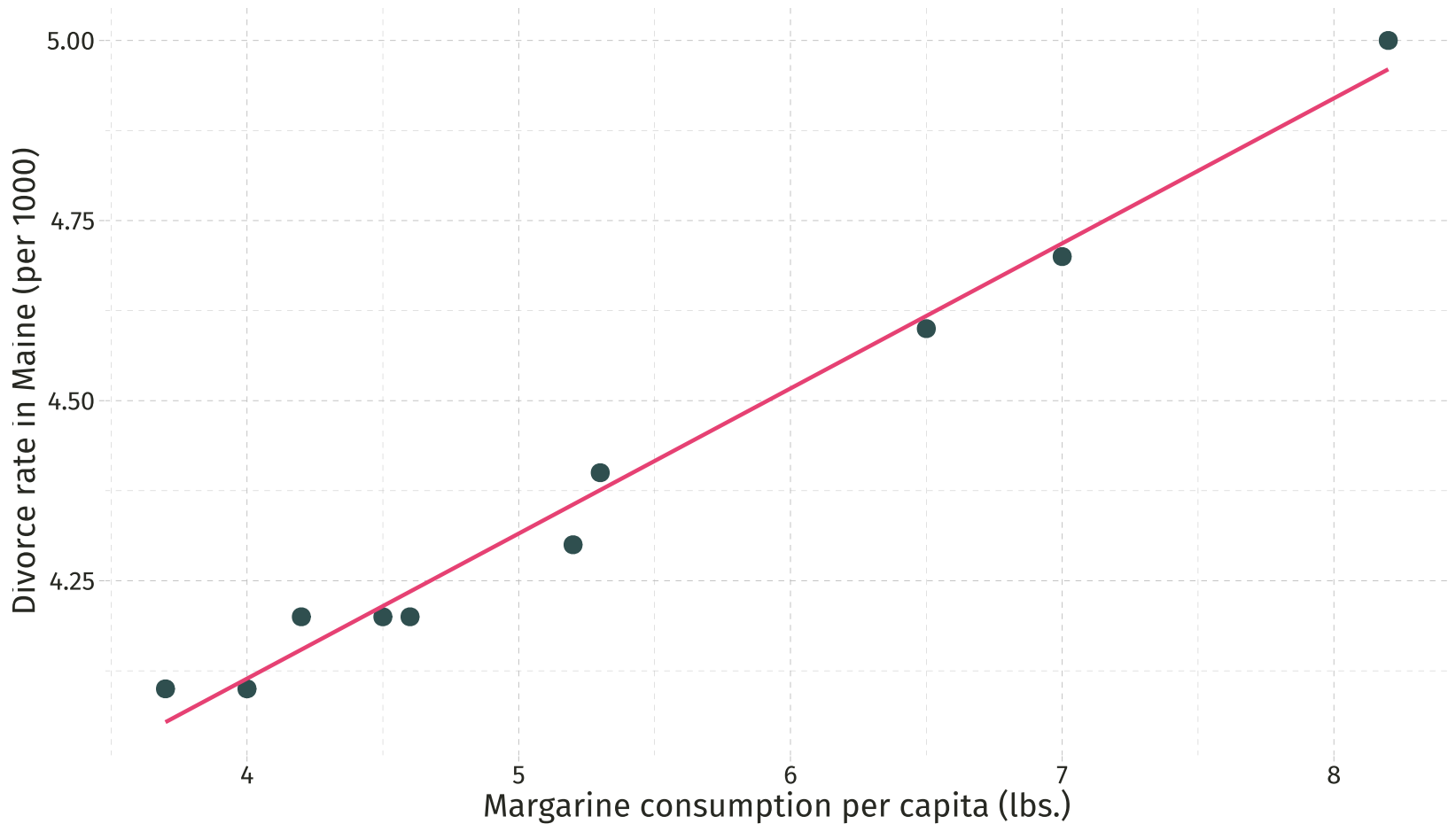
1. **Prediction:** Accurately predict or forecast an outcome given a set of predictors. Given what we know about x , what values do we expect y to take?
2. **Causal identification:** Estimate the effect of an intervention on an outcome. How does y change when we change x ?

The main focus of ECON 4651 and QAMO 3030 is causal identification.

- But...both rely on a common set of statistical techniques.

Econometrics

Not all relationships are causal



Econometrics

Correlation vs. Causation

Common refrain: *"Correlation doesn't necessarily imply causation!"*

- **Q:** Why might correlation fail to describe a causal relationship?
- **A:** Omitted-variables bias, selection bias, simultaneity, reverse causality.

Correlation can imply causation.

- Requires strong assumptions.
- **Real life often violates these assumptions!**
- **Solutions:** Conduct an experiment or find a natural experiment.

Example: *Blue Paradox*

Recent study by economist Grant McDermott and coauthors.

Question: Do commercial fishers preempt fishing bans by increasing their fishing effort before the bans go into effect?

Motivation

- Recent conservation efforts seek to preserve aquatic habitat and increase fish stocks.
- Policy lever: Restrict fishing activity in marine protected areas.
- Concern: Preemptive behavior could *decrease* fish stocks.

Data

- Vessel-level data on fishing effort/intensity.

Example: *Blue Paradox*

Natural Experiment

Phoenix Islands Protected Area (PIPA)

- First mentioned on 1 September 2014; implemented 1 January 2015.
- *Treatment group*: PIPA.
- *Control group*: Outlying Kiribati islands.

Example: *Blue Paradox*

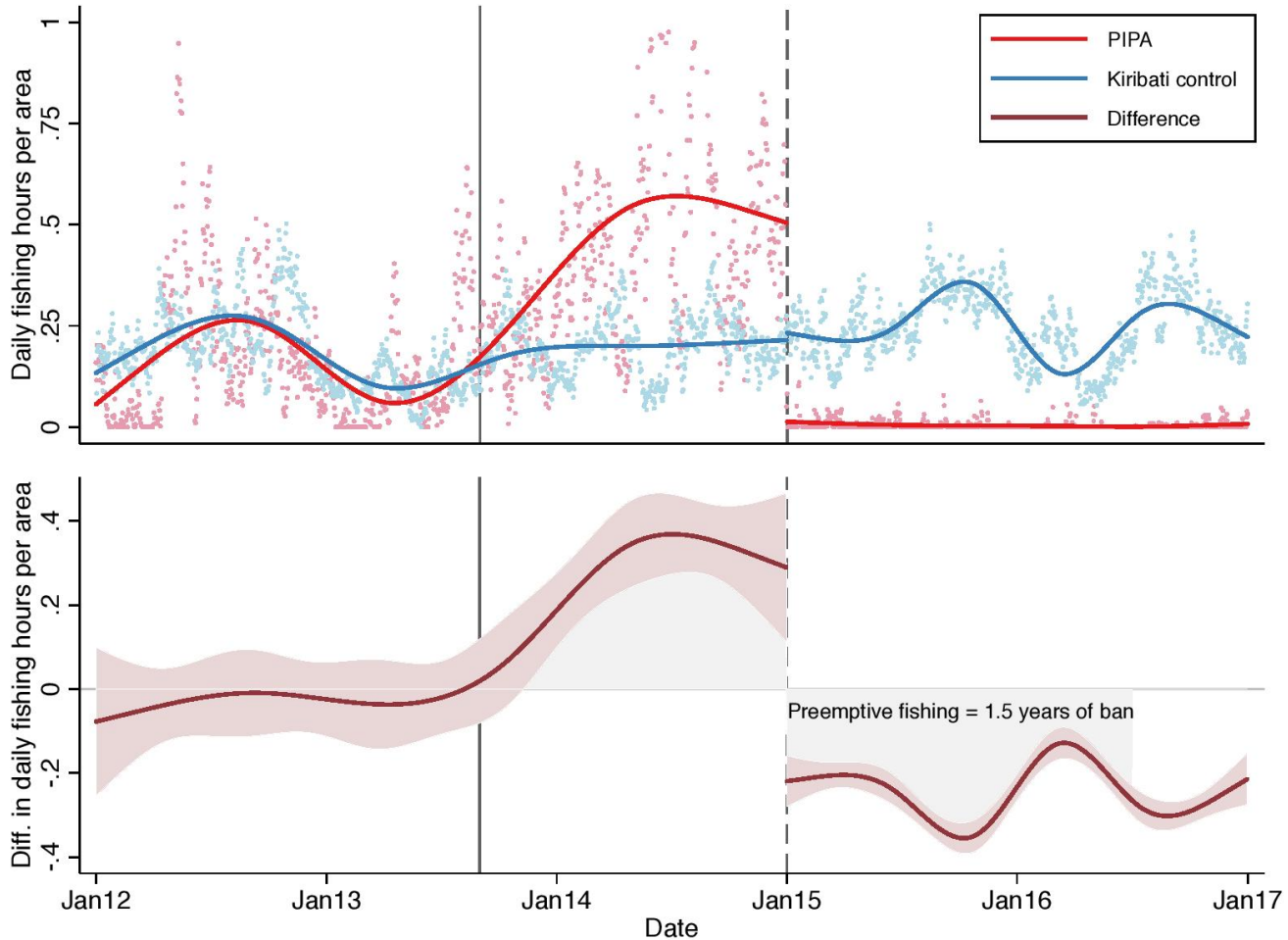
Natural Experiment

Measure the causal effect of the fishing ban by comparing fishing effort in treatment and control regions, before-and-after PIPA.

- A *difference-in-differences* comparison.
- **Assumption:** *Parallel trends*. If we believe this assumption, then the observed change supports a causal interpretation. If not, then the change could reflect other factors and thus fail to isolate the causal effect of the ban.

Example: *Blue Paradox*

Results



Example: *Blue Paradox*

Discussion

Results provide causal evidence that commercial fishers engage in preemptive behavior in response to conservation policy changes.

Results are *consistent* with economic theory, but *cannot prove* that the theory is correct.

- **Science cannot prove anything.**
- Science can **falsify or reject** existing hypotheses or **corroborate** existing evidence.

Also...the causal statement rests on a critical assumption.

- Cannot prove that the assumption is true, but can falsify it.
- Failure to falsify \neq assumption is true.

Econometrics

An applied econometrician[†] needs a solid grasp on (at least) three areas:

1. The **theory** underlying econometrics (assumptions, results, strengths, weaknesses).
2. How to **apply theoretical methods** to actual data.
3. Efficient methods for **working with data**—cleaning, aggregating, joining, visualizing.

This course aims to deepen your knowledge in each of these three areas.

- 1: As before.
- 2–3: **Stata**

[†]: *Applied econometrician* = Practitioner of econometrics, e.g., analyst, consultant, data scientist.

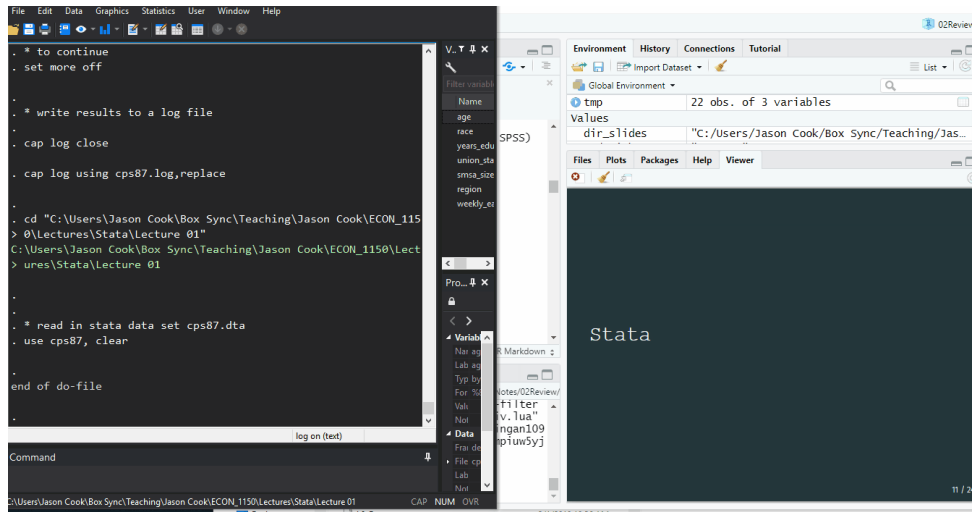


Stata

- There are several statistical packages useful for econometrics, we will use Stata
- Concepts learned easily translate to other packages (e.g., R, SAS, SPSS)
- **Useful Resources:**
 - [Stata FAQ](#)
 - [The Stata listserv](#)
 - [UCLA's resources for learning Stata](#)
- **In Stata:** CPS data
 - Navigating Stata
 - Help files
 - Do-files

Stata - Do File

- Everything you **do** in Stata should be recorded in a **do file**
- The do file should:
 - Begin by opening raw data (source data unedited by user)
 - Perform data cleaning
 - Run analyses and save output
- **Never save over raw data**
- Comment by starting lines with * to organize code



Opening a do file

Stata - help

- Use the help feature liberally
 - Explains what a command does
 - At the end gives examples of using the syntax
 - Sometimes even has video explanations

The screenshot displays the Stata software interface. On the left, the 'Command' window shows a do-file being executed, with the following commands visible:

```
. * to continue  
. set more off  
  
. * write results to a log file  
  
. cap log close  
  
. cap log using cps87.log,replace  
  
. cd "C:\Users\Jason Cook\Box Sync\Teaching\Jason Cook\ECON_1150\Lectures\Stata\Lecture 01"  
> 0\Lectures\Stata\Lecture 01  
C:\Users\Jason Cook\Box Sync\Teaching\Jason Cook\ECON_1150\Lectures\Stata\Lecture 01  
> ures\Stata\Lecture 01  
  
. * read in stata data set cps87.dta  
. use cps87, clear  
  
. end of do-file
```

On the right, the 'Variables' window shows a list of variables with their labels and types:

Name	Label	Type
age	age in years	byte
race	=1 if white	byte
years_educ	years of com	byte
union_status	=1 if in union	byte
smsa_size	=1 if largest	byte
region	=1 if northeast	byte
weekly_earn	usual weekly	float

Below the 'Variables' window, the 'Properties' window shows the 'Data' tab selected, displaying the following information:

Frame	default
Filename	cps87.dta
Label	
Notes	
Variables	7

On the far right, the 'Untitled1.do' editor shows a do-file with the following commands:

```
1 * set it such that the computer does not  
2 * need the operator to hit the return key  
3 * to continue  
4 set more off  
5  
6 * write results to a log file  
7  
8 cap log close  
9 cap log using cps87.log,replace  
10  
11 cd "C:\Users\Jason Cook\Box Sync\Teaching\Jason Cook\ECON_1150\Lectures\Stata\Lecture 01"  
12  
13  
14 * read in stata data set cps87.dta  
15 use cps87, clear  
16  
17 * describe what is in the data set  
18 describe  
19  
20 * generate new variables  
21 * lines 1-2 illustrate basic math functions  
22 * line 3 line illustrates a logical operator  
23 * line 4 illustrate the OR statement  
24 * line 5 illustrates the AND statement  
25  
26 gen age2=age*age //age^2  
27 gen ln_weekly_earn=ln(weekly_earn)  
28 gen union=union_status==1  
29 gen nonwhite=((race==2))((race==3))  
30 gen big_ne=((region==1)&(smsa==1))  
31  
32 label var age2 "age squared"  
33 label var ln_weekly_earn "log earnings per week"  
34 label var union "1=in union, 0 otherwise"  
35 label var nonwhite "1=nonwhite, 0=white"  
36 label var big_ne "1= live in big smsa from northeast, 0=otherwise"  
37  
38
```

Looking up help file for summarize command

Stata - Basic Operations

- **Generate new variables:** `gen age2=age*age`
 - Create a variable (gen) called `age2` and assign (=) to it the value of `age` times `age` (aka age^2)
- **Natural Log:** `gen ln_weekly_earn=ln(weekly_earn)`
- **Binary Variable:** `gen union=union_status==1*`
 - Create a variable called `union` that equals 1 if `union_status` equals 1, otherwise assign `union_status` to be 0

[*] = "assign"; = "compare"

Stata - Basic Operations

- **Logical Conditions (OR):** `gen nonwhite=race==2 | race==3`
 - Create a variable called `nonwhite` that equals 1 if `race` equals 2 *OR* (`|`) 3, otherwise assign `nonwhite` to be 0
- **Logical Conditions (AND):** `gen big_ne=region==1 & smsa==1`
 - Create a variable called `big_ne` that equals 1 if `region` equals 1 *AND* (`&`) `smsa` equals 1, otherwise assign `big_ne` to be zero

Stata - Summary Statistics

`summarize, detail`

- Provides summary statistics of given variable(s), (here, `age`)
- `, detail` option that tells Stata to provide additional info like quantiles

	Age	Education
Mean	37.97	13.16
Median	36	12
Variance	124.4	7.81

Stata - Summary Stats by Categories

```
bysort race: sum weekly_earn
```

- Runs `sum` command separately for each value of `race`
- Here `race=1` is white, `race=2` is black, and `race=3` is Hispanic

Race	Mean Earnings
White	506.5
Black	383.1
Hispanic	368.6

Stata - if Statements

- Add `if` conditions after command to run on subset of data

The screenshot displays the Stata software interface with three main panels:

- Top Panel (Table):** A table showing the distribution of SMSA (Standard Metropolitan Statistical Area) categories. The columns are labeled 1, 2, 3, and Total. The rows represent different regions: =1 if northeast, =2 if midwest, =3 if south, =4 if west. The data is as follows:

	1	2	3	Total
=1 if northeast,	2,806	1,349	842	4,997
=2 if midwest,	56.15	27.00	16.85	100.00
=3 if south, =4 if west	38.46	18.89	15.39	25.10
=1 if largest 19 smsa, =2 if other smsa, =3 not in smsa	1,501	1,742	1,592	4,835
	31.04	36.03	32.93	100.00
	20.58	24.40	29.10	24.29
	1,501	2,542	1,904	5,947
	25.24	42.74	32.02	100.00

Command Window: Shows the command `log on (text)`.

Do-file Editor: Contains the following Stata commands:

```
52 sum weekly_earn age, detail
53
54 * to get means across different subgroups
55 * sample, first sort the data, then gener
56 * summary statistics by subgroup
57
58 sort race
59 by race: sum weekly_earn
60
61 * get weekly earnings for only those with
62 * high school education
63 sum weekly_earn if years_educ>=12
64
65 * get frequencies of discrete variables
66 tabulate race
67
68 * get two-way table of frequencies
69 tabulate region smsa, row column
70
71 *Histogram of age and plot line at averag
72 sum age,d
73 histogram age, normal xtitle(Age) xline(`
74 graph export age_hist.png,replace as(png)
75
76 *****
77
78
79
80
81 * one-sample ttest
```

Variables Panel: Lists variables: Name, age, race, years_educ, union_status, smsa_size, region, weekly_earn.

Properties Panel: Shows properties for the variable 'age': Name, age, Label, age in years, Type, byte, Format, %8.0g, Value label, Notes.

Data Panel: Shows the data file 'cps87.dta' with the following properties: Frame, default; Filename, cps87.dta; Location, C:\Users\Jason Cook\Box Sync\Teaching\Jason Cook\ECON_1150\Lectures\Stata\Lecture 01.

Stata - tabulate Command

- Provides counts of observations in each unique value of variable

The screenshot displays the Stata software interface. The main window shows the results of the `tabulate` command, which has been used to calculate summary statistics for the `weekly_earn` variable across different race categories. The results are presented in three tables, one for each race value (2, 3, and 4).

Table 1 (Race = 2):

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	17,103	506.4874	237.2567	60	999

Table 2 (Race = 3):

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,642	383.095	196.2224	90	999

Table 3 (Race = 4):

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,161	368.5512	200.6758	66	999

The Command window shows the following commands:

```
-> race = 2  
-> race = 3  
-> race = 4
```

The Variables window shows the list of variables in the dataset, including `age`, `years_educ`, `union_status`, `smsa_size`, `region`, and `weekly_earn`.

The Properties window shows the properties of the selected variable, `age`, including its label, type, format, and value labels.

The Command window also shows the following commands:

```
sum weekly_earn age, detail  
* to get means across different subgroups  
* sample, first sort the data, then generate  
* summary statistics by subgroup  
sort race  
by race: sum weekly_earn  
* get weekly earnings for only those with  
* high school education  
sum weekly_earn if years_educ>=12  
* get frequencies of discrete variables  
tabulate race  
* get two-way table of frequencies  
tabulate region smsa, row column  
*Histogram of age and plot line at average  
sum age,d  
histogram age, normal xtitle(Age) xline(`'  
graph export age_hist.png,replace as(png)  
*****  
* one-sample ttest
```

Stata - Figures

- Easiest way to make figures in Stata is using GUI in menu under Graphics → Twoway Graph

The screenshot displays the Stata software interface with three main panels:

- Command Window (Left):** Shows the execution of commands to filter data by race and calculate summary statistics for weekly earnings.
- Variable Explorer (Middle):** Lists available variables and their properties.
- Do-file Editor (Right):** Contains a script of Stata commands for data analysis and visualization.

Command Window Output:

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,642	383.095	196.2224	90	999

→ race = 3

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,161	368.5512	200.6758	66	999

. sum weekly_earn if race==3

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,161	368.5512	200.6758	66	999

log on (text)

Command

C:\Users\Jason Cook\Box Sync\Teaching\Jason Cook\ECON_1150\Lectures\Stata\Lecture 01

Variable Explorer:

Name	Label	Type	Format	Value label	Notes
age	age in years	byte	%8.0g		
race		=1			
years_educ	years of education	byte	%8.0g		
union_status	union status	=1			
smsa_size	size of standard metropolitan statistical area	=1			
region	region	=1			
weekly_earn	weekly earnings	ust			

Do-file Editor (cps87):

```
52 sum weekly_earn age, detail
53
54 * to get means across different subgroups
55 * sample, first sort the data, then generate
56 * summary statistics by subgroup
57
58 sort race
59 by race: sum weekly_earn
60
61 * get weekly earnings for only those with
62 * high school education
63 sum weekly_earn if years_educ>=12
64
65 * get frequencies of discrete variables
66 tabulate race
67
68 * get two-way table of frequencies
69 tabulate region smsa, row column
70
71 *Histogram of age and plot line at average
72 sum age,d
73 histogram age, normal xtitle(Age) xline(`
74 graph export age_hist.png,replace as(png)
75
76 *****
77
78
79
80
81 * one-sample ttest
```

Line: 74, Col: 15 CAP NUM OVR

Stata - Documenting code

- Organize code into sections
- Use `*` to comment entire lines of code or `\\` to comment out everything following it on the same line
- This can be used to write notes about what the chunk of code is doing

Example

```
*Create new variables
```

```
gen age2 = age^2 \\other comment here
```

In the example above, only the `gen age2=age^2` part of the code is executed, Stata will skip over the rest of the text

Questions?