# ECON300 Econometrics

## Lecture 1 Introduction

# Outline

- Course information
- Introduction to econometrics
- Causal question examples
- Structure of data

# Basic information

- **Lecture Session:**
  -**Location**: Burnham Hall Room 317
  -**Time**:  Monday & Wedneaday 4:30-5:45 pm

- **Instructor**: Salman Khan
  - **Office**:  UH819
  - **E-mail:** skhan275@uic.edu
  - **Office Hours**: Tuesday 11:15- 12:15p

- **Teaching Assistant**: Taisiia Stanishevska
  - **Office**:  UH813
  - **E-mail:** : tstani4@uic.edu
  - **Office hour:** Thursday 1-2pm

# Access to STATA

- Option 1: Use lab computers- Stata 14 is now installed in **SCE401, LIB IDEA Commons, and BSB B001.** Contact ACCC if you have questions about STATA availability.

- Option 2 (requires Internet access at all times):  One option requires internet access to login to a remote server. A license for the semester is $25 from the UIC web store.
https://webstore.illinois.edu/shop/product.aspx?zpid=3777

- Option 3 (does not require internet access):  An alternative is to download & install a copy that you can buy directly from Stata for $48 for a 6-month license.
https://www.stata.com/order/new/edu/gradplans/student-pricing/

# STATA (useful resources)

Several resources for learning to use Stata:

- Official: http://www.stata.com/support/documentation/
- Tutorials: http://www.ats.ucla.edu/stat/stata/modules/
- Forum: http://www.statalist.org/forums/

# Course requirements

Total points:1000
- Exams:600
- Problem Sets: 300
- Quizzes:100

# Course requirements

- Exams (600 pts)
  - Exam 1 (200 points) is scheduled for **,Wednesday, Feb 19th**
  - Exam 2 (200 points) is scheduled for **Wednesday, Mar 19th**
  - A **cumulative** final exam (200 points) will take place during the final week

- Problem Sets (50*6=3000 pts)
  - The problem sets cover materials from lecture sessions and book.
  - There are 6 problem sets in total; May give a extra credit problem set
  - **Only hard copies are accepted**
  - **No late homework is accepted**

# Course requirements

- Quizzes (100 pts)
  - Given randomly at the beginning of regular lecture.
  - Serve as review and attendance tool
  - No makeup quizzes

# Grading

Grades will be based on the following distribution:

A    900 – 1000 points
B    800 – 899 points
C    700 – 799 points
D    600 – 699 points
F    0 – 599 points

# Textbook

- Textbook:

    *Real Econometrics: The Right Tools to Answer Important Questions* (ISBN: 9780190296827)


- Recommended Textbook:

    Josh Angrist & Jörn-Steffen Pischke, *Mastering Metrics*, (ISBN 9780691152844)

# Some tips about learning Econometrics

- Come to class and participate.
- Most returns to understanding and asking questions in class.
- Review notes for pop quizzes before each lecture
- Focus on the study guides before actual exams.
- For STATA questions, don't hesitate to contact me or TA for help if you get stuck with coding
- Talk to TA/me/peers for any learning difficulties/feedback

# What is econometrics?

- "Economists' use of data to answer cause-and-effect questions; Analyze and interpret data using statistical inference"-*Mastering 'Metrics*

- "Statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing policy"-*Introductory Econometrics A modern Approach*

- Prerequisites: Econ270 Statistics for economics or equivalent

# Course goals

- Gain familiarity with common econometric techniques (the Furious Five)
    - Regression
    - Random assignment
    - Regression discontinuity
    - Differences-in-differences
- Become capable of evaluating public policies/non-technical articles
- Develop data analyzing skills using STATA

# Course structure

- Statistics Review and Regression
  - Foundation
  - Week 1-6

- Cause and effects techniques
  - Advanced micro-econometrics
  - Week 1 and Week 7-15

# Example 1- Return to selective colleges

- Does going to Harvard University increase one's earning?
- https://www.youtube.com/watch?v=iPBV3BlV7jk
- Naïve comparison: compare earnings of people who went to Harvard with those went to UMass? Is it a trust-worthy comparison? Why?
- What is the ideal comparison?
- Ideal comparison: compare earnings of a person who went to Harvard with the same person if he had gone
    - -Mission impossible!

# Example 2-Health Insurance

- The Affordable Care Act (ACA)
  https://www.youtube.com/watch?v=C2nskzMgZ_g

- Is it a good thing? What is your opinion?

  -Cost vs. benefit? From a social planner perspective

- One important question on the benefit side: does health insurance make people healthier?

  -A big research question in the field of health economics

  -How would you approach this question?

# Mean comparison:

- Naïve comparison: compare the health status of people who have health insurance with those who have not. Is this a good comparison? Why?

  -People with insurance are NOT the same as people without insurance!

  -Jargon: we are not comparing **apple to apple**, or the uninsured are NOT good **counterfactuals** of the insured.

- What is the ideal comparison?

  -Compare the health status of the same person at the same point in time with and without insurance, other things equal (ceteris paribus)

  -This is never possible

  -"The Road not taken"

# Exercise

- Obesity: https://www.youtube.com/watch?v=IkoC3HZwe7Q
- Research Question: does fast foods make people heavier?
- To make it simple, assume there are two types of people: who eat only fast foods, and who never eat fast foods
- What is the naive comparison? Why is it misleading?
- What is the ideal comparison?

# Two types of "econometrics"

- "Model based":

  -Apply multiple models to fit data; identify the best model according to certain criteria, e.g. forecasting, machine learning

- "Design based":

  -Most applied micro-econometrics belong here – focus on research designs: finding credible answers to causal questions

- This course focuses on the second type

# Structure of data

- Common types of data:

  -Cross-sectional data

  -Time series data

  -Pooled cross-sectional data

  -Panel or longitudinal data

# Cross-sectional data

- Consists of a sample of individuals (households, firms, cities, etc.) taken at a **given point in time**

<span style="color:red">variables (column)</span>

<span style="color:red">observations (row) →</span>

<span style="color:red">Note: year doesn't vary, ID varies</span>

| ID | year | wage | education | female | married |
|----|------|------|-----------|--------|---------|
| 1 | 1998 | 3.10 | 11 | 1 | 0 |
| 2 | 1998 | 3.24 | 12 | 1 | 1 |
| 3 | 1998 | 3.00 | 11 | 0 | 0 |
| 4 | 1998 | 6.00 | 8 | 0 | 1 |
| 5 | 1998 | 5.30 | 12 | 0 | 1 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| 525 | 1998 | 11.56 | 16 | 0 | 1 |
| 526 | 1998 | 3.50 | 14 | 1 | 0 |

# More about categorical variable

- As a general rule, we convert categorical variables into numerical format

- Example: education level has three categories-high school dropout; high school grad and some college; college grad and above. We code it as 0 for HS dropout, 1 for HS grad and 2 for college grad

- Specifically, if a categorical variable takes on only two values such as gender, whether age 18 or above, etc. We code it as a 0, 1 variable, and call it a **dummy variable**

- Once coding a category as 0, we can think of it as a baseline category

# Time series data

- Consists of the same observation **over time**; e.g. stock prices, annual GDP, annual homicide rates, etc.; the frequency that data is collected can be daily, weekly, monthly, yearly, etc.

Note: ID doesn't vary, year varies

| ID | year | unemployment | GNP |
|----|------|--------------|--------|
| 1 | 1950 | 15.4 | 878.7 |
| 1 | 1951 | 16.0 | 925.0 |
| 1 | 1952 | 14.8 | 1015.9 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 1 | 1986 | 18.9 | 4281.6 |
| 1 | 1987 | 16.8 | 4496.7 |

# Pooled cross-sectional data

- A data set with both cross-sectional and time series features e.g. combining two cross-sectional data sets of U.S. household surveys, one taken in 1993 and one in 1995

Note: Both ID and year vary, but the 1995 observations are NOT the same as in 1993

| ID | year | house price | square feet | bedrooms |
|-----|------|-------------|-------------|----------|
| 1 | 1993 | 85,500 | 1,600 | 3 |
| 2 | 1993 | 67,300 | 1,440 | 3 |
| . | . | . | . | . |
| . | . | . | . | . |
| 251 | 1993 | 243,600 | 2,600 | 4 |
| 252 | 1995 | 65,000 | 1,250 | 2 |
| . | . | . | . | . |
| . | . | . | . | . |
| 520 | 1995 | 57,200 | 1,100 | 2 |

# Panel or longitudinal data

- Consists of time series for each cross-sectional member of the data set; e.g. city crime history for a set of cities followed over a twenty-year period

<span style="color:red">-Note: Both ID and year vary, and the 1990 observations are the same as in 1986.
-This is the key difference between Panel data and pooled cross-sectional data!</span>

| City ID | year | murders | population |
|---------|------|---------|------------|
| 1 | 1986 | 5 | 350,000 |
| 1 | 1990 | 8 | 359,200 |
| 2 | 1986 | 2 | 64,300 |
| 2 | 1990 | 1 | 65,100 |
| . | . | . | . |
| . | . | . | . |
| 149 | 1986 | 10 | 260,700 |
| 149 | 1990 | 6 | 245,000 |
| 150 | 1986 | 25 | 543,000 |
| 150 | 1990 | 32 | 546,200 |

# Exercise

Which type of data are the followings?

- US GDP growth:
  http://www.multpl.com/us-gdp-growth-rate/table/by-year

- Illinois high school ranking:
  http://www.schooldigger.com/go/IL/schoolrank.aspx?level=3

- NBA champions:
  https://www.ticketcity.com/nba/nba-finals-tickets/nba-finals-champions.html

- Current Population Survey (CPS)

# Review for quizzes:

- Be able to translate a problem into mean comparison
- Why naïve comparison is misleading
- What is the ideal comparison
- Be able to recognize four types of data