ECON 1190: Econometrics 2:
Class 1: Intro and Stats Review Blitz

Claire Duquennois

Syllabus and course logistics

Course description

- hands-on skills in applied econometric research methods
- focus on research design and approaches to establish empirical relationships of cause and effect
- advanced methods in regression analysis

Topics covered include:

- omitted variable bias, fixed effects, instrumental variables, randomized control trials, difference-and-differences, and regression discontinuities.
- learning the basic tools of programming and coding in R (no prior knowledge is required)
- learn to present results in nice formatted tables and figures

Is this course right for me?

- I am interested in how economists do research with data
- I am interested in developing my (highly marketable!) data analysis skills
- I have taken a class such as ECON 1150: Applied Econometrics 1 (or STAT 1221, IF 1071) and am comfortable with:
  - basic probability and statistics
  - basic linear regressions
- to use other classes as prerequisites, check with Katherine Wolfe (kwolfeATpitt.edu)
- I am ready to work hard!

Contact

Email: ced87-at-pitt.edu

Office hours: Thursdays after class 5:15-6:15 in 4912 Posvar

Personal advising/discussion: Sign up at https://sites.google.com/view/claireduquennois/teaching

Required materials

- Rstudio Cloud Student Account:
  - Purchase a Student account ($5/month for duration of term).
- Homework article readings:
  - 5 academic journal articles that we will replicate
  - available on canvas to be read by the dates indicated in the course schedule
- Lecture notes and slides available on course github page: https://github.com/claireduq/econ1190
- Reference texts as needed:
  - Nick Huntington-Klein The Effect (Available online).
  - Cunningham, Scott. Causal Inference: The Mixtape (Available online)
  - Wooldridge, Jeffery. Introductory Econometrics: A modern Approach. (Any edition).
  - Angrist, Joshua D and Pischke, J. Mostly Harmless Econometrics. Princeton University Press, 2008.

Grades

- Homework assignments 50%: 6 assignments, best 4 count for 10% each, lowest 2 for 5% each)
- 2 Exams 40%: Your best performance is worth 25% your lowest 15%
- 26 daily top hat 'Tiny Quizzes' 10%:
  - Participation 6%: Each of 20 for 0.3% (6 lowest are dropped)
  - Accuracy 4%: let score be the number of correct answers on tiny quizzes
    - if $8 <$ score ≤ 11: 1%
    - if $11 <$score ≤ 14: 2%
    - if $14 <$ score ≤ 17: 3%
    - if $17 <$ score: 4%

Homework assignments

- To deepen your understanding, practice coding in R, and see how real research is done in empirical applied economics
- To be done in R markdown via R Studio Cloud, submitted as PDF's on Gradescope
- You can help each other but submitted assignments cannot be identical (0 grade).
- submit by due date (-5% for each day late)
- Regrade request to be done within a week of getting graded assignment. Avoid regrade requests for minor point changes.

Exams

- will be written, timed, traditional pen and paper exams
- you will not be asked to code but will be asked to interpret code
- you will be allowed 1 page of handwritten notes
- see MQE exam format to get an idea of what to expect

Other first day stuff

- Introduce yourself!

Send me an email and let me know if there is anything I need to know or you would like to share about yourself (eg. unique interests, pronouns, preferred names, special accommodations, anticipated absences on exams. . .) reach out and keep in touch!

Review Blitz: Probability, Statistics and regressions

Review: Probability

The probability of event A ($P(A)$) is a number between 0 and 1:

$$0 \leq P(A) \leq 1$$

Conditional Probability

The probability of an an event might change when additional information becomes available

$$P(A|B) = \text{probability of A given that B occured}$$

Statistical Independence

Two events are statistically independent if the information that one of them has happened does not change the probability of the other

$$\text{A and B are independent if } P(A|B) = P(A)$$

Stats Review: Random Variables

Random variables:

a variable that whose value depends on random events

For a given outcome x we may be interested in the probability that the variable X achieves that outcome

$$P(X = x) = ?$$

These outcomes could be discreet or continuous

Stats Review: Discreet random variables

Discreet random variables take a finite number of values

There is a probability associated with each of it's possible values

$$\sum_{allx} P(X = x) = 1$$

Stats Review: Continuous random variables

Continuous random variables take an infinite number of possible values

The density function gives the probability that a random variable is smaller than some value so that

$$P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

with

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

A cumulative density function maps out

$$F(a) = P(X \leq a)$$

Stats Review: Summarizing random variables

Measures of location:

- mean:

$$\text{Discreet: } E[X] = \mu_X = \sum_{allx} xP(X = x)$$

$$\text{Continuous: } E[X] = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$$

- Also median, percentiles (ex: quartiles) ...

Stats Review: Summarizing random variables

Measures of dispersion:

- Variance

$$\text{Discreet: } Var(X) = \sigma_X^2 = \sum_{allx} (x - \mu_X)^2 P(X = x)$$

$$\text{Continuous: } Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx$$

- Standard deviation:

$$Std.Dev.(X) = \sigma_x = \sqrt{\sigma_X^2}$$

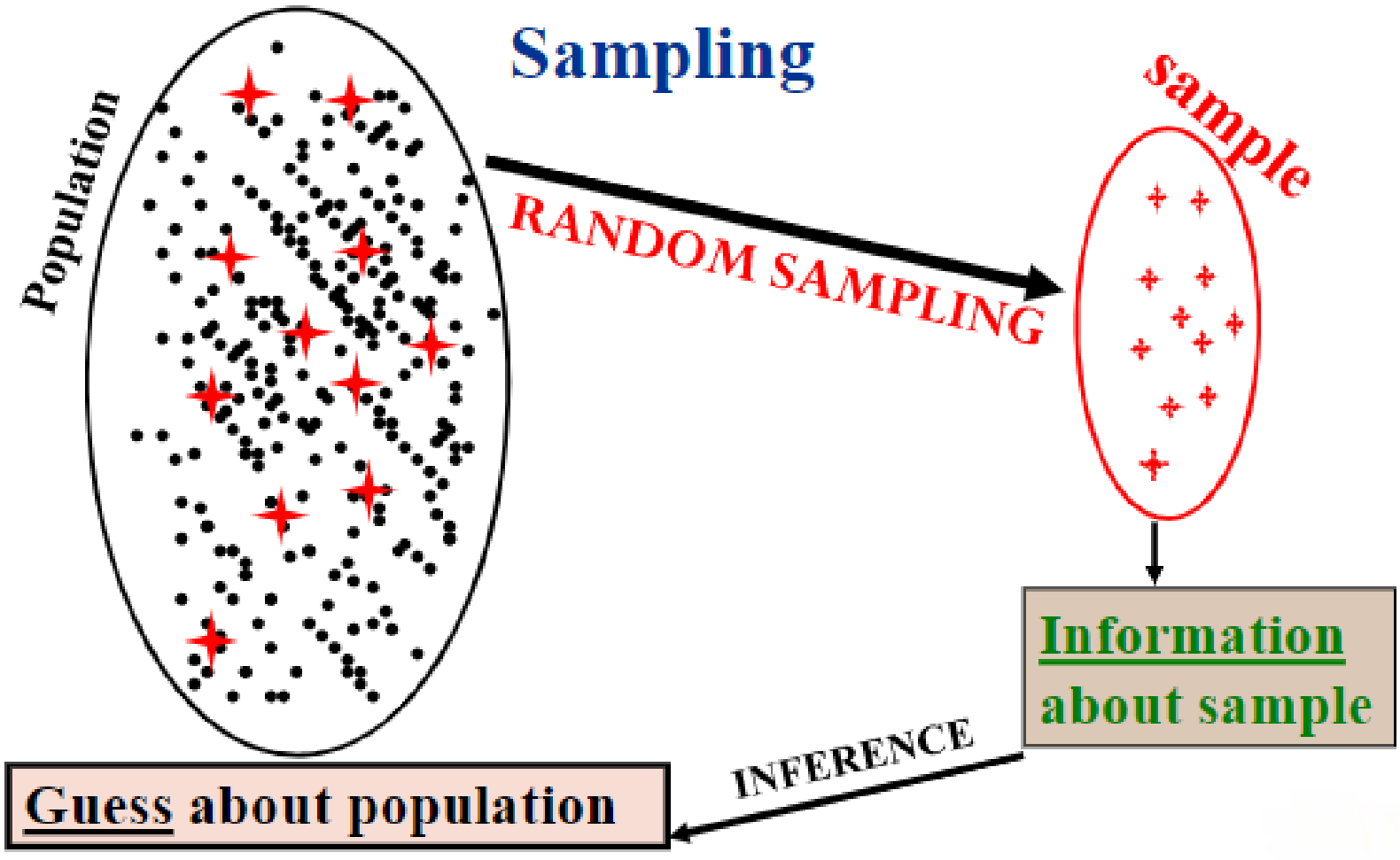
Stats Review: Important distributions

Uniform: the same probability to all values of the random variable

Normal: Important mainly because of the central limit theorem

Standard normal: A normal with mean 0 and standard deviation 1

Stats Review: Sampling, estimation and inference



Stats Review: Sampling, estimation and inference

What information do we get from the sample?

Get the sample mean

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

If you collected a new random sample, your sample mean will be different:

$$\bar{X}_1 \neq \bar{X}_2 \neq \bar{X}_3 \neq \bar{X}_4 ...$$

So what can we learn from the SAMPLE mean about the POPULATION mean μ (what we really care about)?

Stats Review: The Central Limit Theorem

Consider a random sample of n observations taken from a population X with expected value μ_X and standard deviation σ_X. The following are always true of the sample mean $\bar{X}$.

- The probability distribution of $\bar{X}$ approaches a normal distribution as n increases (i.e. with bigger sample size, the distribution of possible $\bar{X}_i$'s looks like a bell curve)
- The expected value of $\bar{X}$ equal μ_X (i.e this bell curve is centered on the population mean)
- The standard deviation of $\bar{X}$ (aka the **standard error**) equals $\frac{\sigma_X}{\sqrt{n}}$ (i.e. the bell curve is tighter as the sample size increases)

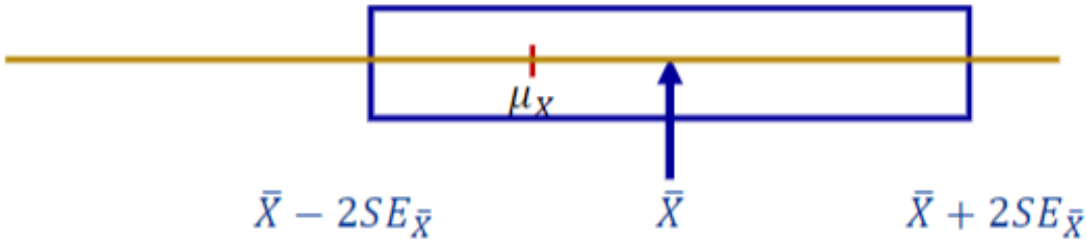
Stats Review: Standard errors and confidence intervals

So given our sample, our best guess of the population mean is the sample mean

$$E[\mu_X] = \bar{X}$$

How good is this guess? How far away from the population mean from the sample mean?

- We build confidence intervals around $\bar{X}$ with its standard error $SE_{\bar{X}}$
- To do this we use a critical value, often $1.96 (\approx 2)$, that reflects how certain we want to be that the confidence interval includes the population mean.



Stats Review: Hypothesis testing

Using these confidence intervals, we can REJECT or FAIL TO REJECT a Null Hypothesis.

- Reject the Null Hypothesis if there is "enough' ' evidence against it

p-value: The largest probability of observing sample data as extreme as your draw given that the Null is true

Upcoming

Next class:

- Intro to R Studio Cloud and R Markdown
- (consider bringing your laptops)
- To do before class: Sign up for a student account ($5 / month) and accept my invitation to the class workspace

Tuesday:

- More review: Regressions and regressions in R