UC San Diego

ECON 5/POLI 5D: Data Analytics for the Social Sciences

Course Syllabus
Winter 2022

Instructor	**Jiannan Zhao** (jiz717@ucsd.edu) Office Hours: Monday 10-noon SSB 351 (Social Sciences Building) or Zoom Zoom Link: [redacted] Office hour sign-up link Here
TA	**Bert Wilden** (bwilden@ucsd.edu) Office hours: Monday 3-5pm SSB 347 (Social Sciences Building) or Zoom Zoom Link: [redacted] Office hour sign-up link Here

Course Description

As data about individuals, organizations, and governments become increasingly available, social data analytics are transforming the way we think about the economy, politics, and society. This course will teach skills necessary to navigate the world of social data. We will learn basic principles of coding through the lens of Excel, Stata, and R. While learning coding fundamentals, we will shed light on basic research design principles and statistical concepts that will help us grasp social questions that the era of a society governed by data presents us.

Weeks 1-2: Remote instruction

Remote classes and labs	MWF 9-9:50am Zoom link (same for classes and labs) Passcode: econpoli5

Weeks 3-10: In-person instruction

In-person classes	MW 9-9:50am Pepper Canyon Hall 280
In-person labs	F 9-9:50am Sequoyah Hall 142

Learning Outcomes

After completing this course, students will be able to:

1. Demonstrate competency in the logic of research design.
2. Demonstrate competency in the manipulation and analysis of data sets.
3. Employ research design and data analysis skills to better understand political, economic, and social relationships.

Course Format

On weeks 1 and 2, the course will be taught remotely. From weeks 3-10, the course will be taught in person.

Pre-recorded Lectures

Much of the course content will be taught through pre-recorded lectures. They will be posted on Canvas, and students must watch them **before the Wednesday class of each week.** Some of the lecture videos will be designated as "Advanced." This means that the skills taught in these videos are not required to complete the homework assignments and the final project. However, we highly encourage you to learn these skills as they will likely be helpful in your professional career.

Classes

Classes (Mondays and Wednesdays, 9-9:50am) will focus on active learning activities and group discussions to think critically about quantitative research in the social sciences. For the Wednesday classes, students will be expected to watch all the lecture videos up to that week. Classes will be via Zoom on weeks 1 and 2, and in-person at Pepper Canyon Hall 280 from weeks 3 to 10.

Labs

Labs (Fridays, 9-9:50am) will focus on hands-on activities to learn Excel, Stata, and R. You will be given a PDF with a list of tasks to complete during each lab. Labs will be via Zoom on weeks 1 and 2, and in-person at Sequoyah Hall 142 from weeks 3 to 10.

Canvas Page

Pre-recorded lectures, course materials, and assignment submissions will be posted on Canvas. All assignments will be turned in via Canvas:

https://canvas.ucsd.edu/courses/34792

Please be sure to explore and familiarize yourself with the course Canvas page.

Online Discussion Forum

Given the online format of this class, it will be helpful to share questions and answers on a platform that all students may access. In this class, we will use the **Piazza** online discussion board. Piazza is a question-and-answer platform that supports code formatting, embedding images, attaching files, and customized email frequencies.

https://piazza.com/class/kx9dv87akrr4xv

Be sure to use Piazza (instead of emailing the instructional team) when asking questions about course content. A question that you may have can be useful for another student. If you want to remain anonymous, you can hide your name from other students on Piazza.

Assignments, Projects, and Grading

The content of this course is cumulative, and each week builds upon the previous week. It is critical to keep up with the pre-recorded lectures and complete every assignment on time. For assignment, late submissions will lose 5% of the grade for every day late. No submissions more than three days late will be accepted.

Quizzes (10%)

Students will take five quizzes on Canvas that will assess your basic knowledge of the logic of research design. Each quiz will have 5-10 multiple-choice questions on the content taught in recorded lectures of the previous week. Quiz questions will be drawn randomly from a larger bank of questions, which will allow you take the same quiz multiple times—as many times as you want—so long as you complete it before the due date. Quizzes will be graded automatically, and only the highest grade on each quiz will count. Therefore, taking a quiz an additional time can only improve your score.

Homework assignments (40%)

There will be three homework assignments that will serve as periodic checks on your competency in the manipulation and analysis of data sets using Excel, Stata, and R. For each homework, you will submit a PDF with the results of some data analysis task and written interpretations of these results. All homework assignments will be submitted via the Gradescope tab on Canvas. If this is your first time using Gradescope, please watch this video and budget enough time to familiarize yourself with the user interface.

Final project (40%)

Students will complete an independent project that demonstrates mastery of the material taught during the quarter. The project will be due on **March 16 (Wednesday) at 11:59pm** but updates will be due throughout the quarter with homework submissions. Further, every student must schedule a one-on-one meeting with the instructor in weeks 7-8 for a check-in about the state of the project.

Participation (10%)

The participation grade will be based on class and lab attendance, efforts to participate in class activities and discussions, and office hours visits. The instructor will make a holistic assessment across these criteria and assign participation grades in two installments, on weeks 5 and 10.

Summary of Grade Criteria

Assignment	Weight	Due Dates
Quizzes	10%	Sundays (1/9, 1/23, 1/30, 2/13, 2/20) 11:59pm
Homework assignments	40%	Sundays (1/16, 2/6, 2/27) 11:59pm
Final Project	40%	Wednesday (3/16) 11:59pm
Participation	10%	No submission required
	100%	

Grading Procedures and Grade Appeal Policy

Quizzes: Quizzes will be made entirely of multiple-choice questions and they will be graded automatically. All quiz questions will be submitted into canvas by your instructor. Your instructor may make mistakes when submitting questions—a correct answer could be marked as incorrect, and vice-versa. If you believe that this is happening on your quiz, please take a screenshot and email it to me (jiz717@ucsd.edu). I will review it and fix the quiz (and your grade) if needed.

Assignments: All homework assignments will be graded by your TA. The instructor and your TA will split the grading of the final project. The instructional team will be grading many assignments in a relatively short time window, and grading mistakes can happen. If you believe that there is a mistake in how your assignment was graded, please submit a regrade request via the Gradescope tab on Canvas. You can find instructions on viewing feedback and requesting regrades in this video. We will only consider regrade requests submitted within a week of you receiving your score.

Course Materials and Tools

Technology Requirements

We will use three statistical software programs commonly used by social scientists: Excel, Stata, and R. Excel and Stata both require licenses that are available for free to UCSD students. R is open-source and free to everyone. Instructions on how to install the three software packages have been posted on Canvas.

Course Schedule

Each week has pre-recorded lectures on Canvas. **Pre-recorded lectures must be watched before the Wednesday class of each week.** Please note that the syllabus will continue to be updated throughout the quarter.

Week 1: Theories, Hypotheses, and Variables

Learning outcomes:

- Identify a good empirical research question.
- When presented with a hypothesis, identify the dependent variable, independent variable, and unit of analysis.
- Identify the type of a variable.

Monday, January 3: Class 1

- Introduction.

Wednesday, January 5: Class 2

- Discuss the promises and limitations of quantitative analysis.

Friday, January 7: Lab 1

- Install Excel, Stata, and R.

Quiz 1 due on Sunday, January 9 at 11:59pm.

Week 2: Descriptive Analysis in Excel

Learning outcomes:

- Explain the difference between descriptive analysis and inferential analysis.
- Propose ways to describe data given the types of variables.
- Conduct descriptive analysis using Excel.

Monday, January 10: Class 3

- Propose ways to measure a variable.

Wednesday, January 12: Class 4

- Draw bivariate plots.

Friday, January 14: Lab 2

- Conduct descriptive analysis in Excel

Homework 1 due on Sunday, January 16 at 11:59pm.

Week 3: Sampling and Intro to Stata

Learning outcomes:

- Explain the difference between population and sample.

- Explain how the sampling procedure affects the conclusions we can make.
- Conduct descriptive analysis using Stata.

No class on Monday, January 17 (Matin Luther King Jr. Day.)

Wednesday, January 19: Class 5
- Assess sampling procedures.

Friday, January 21: Lab 3
- Conduct descriptive analysis in Stata.

Quiz 2 due on Sunday, January 23 at 11:59pm.

Week 4: Hypothesis Testing and Data Wrangling in Stata

Learning outcomes:
- Assess measurement bias and validity.
- Interpret the result of a hypothesis test.
- Conduct a difference in proportions test using Stata.

Monday, January 24: Class 6
- Discuss different types of sampling bias.

Wednesday, January 26: Class 7
- Assess measurement bias and validity.

Friday, January 28: Lab 4
- Run a difference in proportions test in Stata and interpret the results.

Quiz 3 due on Sunday, January 30 at 11:59pm.

Week 5: Regression Analysis in Stata

Learning outcomes:
- Explain the logic of hypothesis testing with p-values.
- Interpret regression results.
- Run a simple regression in Stata.

Monday, January 31: Class 8
- Regression

Wednesday, February 2: Class 9
- Regression

Friday, February 4: Lab 5
- Run a simple regression in Stata and interpret the results.

Homework 2 due on Sunday, February 6 at 11:59pm.

Week 6: Confounding Variables and Intro to R

Learning outcomes:

- Think of potential confounding variables for a given hypothesis.
- Think of potential intervening variables for a given hypothesis.
- Perform basic operations in R.

Monday, February 7: Class 10

- Think of potential confounding variables for different hypotheses.

Wednesday, February 9: Class 11

- Intro to R

Friday, February 11: Lab 6

- Basic operations in R.

Quiz 4 due on Sunday, February 14 at 11:59pm.

Week 7: Introduction to Causal Inference and Data Wrangling in R

Learning outcomes:

- Explain how random assignment deals with confounding variables.
- Explain how to deal with confounding variables using observational data.
- Data wrangling in R.

Monday, February 14: Class 12

- Intro to causal inference – random assignment

Wednesday, February 16: Class 13

- Selection on observables

Friday, February 18: Lab 7

- Data wrangling in R

Quiz 5 due on Sunday, February 20 at 11:59pm.

Week 8: Multivariate Regression and Data Visualization in R

<u>No class on Monday, February 21 (Presidents' Day.)</u>

Wednesday, February 23: Class 14

- Multivariate regression

Friday, February 25: Lab 8

- Data visualization in R

Homework 3 due on Sunday, February 27 at 11:59pm.

Week 9: Linear Regression in R

Monday, February 28: Class 15

- Linear regression in R

Wednesday, March 2: Class 16

- Linear regression in R

Friday, March 4: Lab 9

- Linear regression in R

Week 10: Dummy Variables, Functions, and R Markdown

Monday, March 7: Class 17

- Dummy variables

Wednesday, March 9: Class 18

- Functions and R Markdown

Friday, March 11: Lab 10

- Functions and R Markdown

Final project due on Wednesday, March 16 at 11:59pm.