

University of California, Los Angeles
Department of Political Science

Introduction to Data Analysis

Political Science 6
2017 Fall Quarter

Syllabus

<i>Meetings:</i>	Dodd 121, M/W 8:00 AM – 9:15 AM
<i>Course Website:</i>	https://moodle2.sscnet.ucla.edu/course/view/17F-POLSCI6-1
<i>Instructor:</i>	Jesse Acevedo
<i>Email:</i>	jesseacevedo@ucla.edu
<i>Office Hours:</i>	Mondays 1-3pm, 4280 Bunche

This course introduces data analytic methods used by social scientists to make inferences about how the world works. The curriculum is designed with 3 objectives in mind: 1) to introduce students to basic statistical theory. Toward this end, we will examine fundamental statistical concepts, the logic of causality, and common analytic challenges (e.g. confounding, bias, correlation versus causation etc.). 2) to acquaint students with some of the tools used by social scientists to surmount said challenges and test hypotheses about the world – contingency tables, t-tests, regression, etc. 3) to teach students how to present their data and results by effectively combining prose and graphical presentations (e.g. tables, histograms, scatterplots, boxplots). In summary, the goal of this course is to teach students how to think about and analyze data, and present their findings in a clear and compelling manner. We will analyze data from many aspects of politics in both developed and developing countries, including data on elections, economic development, indicators of human welfare and the effects of these factors on one another.

Grading

Your grade will be based on three data essays (i.e. homework assignments), a midterm, and a final consisting of two parts – final data project and an in-class final exam.. A total of 100 points will be allocated as follows:

<i>Item</i>	<i>Due Date</i>	<i>Percent of final grade</i>
Data Essay 1	Oct 20	10
Data Essay 2	Nov 3	10
Data Essay 3	Nov 17	10
Midterm	Nov 1	20
Final data project: Research Paper	Dec 15	20 or 30**
Final Exam (In-class Exam)	Dec 13	20 or 30**

The data essay assignments are exercises in which one begins with a dataset, does some statistical analysis on it, presents the results in tables and graphs, and describes and analyzes the results. The data essays will be structured; the final data project will be open-ended. Note that 30

percent of your final grade is based on the 3 homework assignments and so you must take them just as seriously as you would the midterm and final exams. The best way to learn data analysis is to **do** data analysis and the homework assignments are designed to get you to do just that. There is NO extra credit. For grade disputes, please send an email explaining the parts of the exam that are in question and these will be the only parts of the exam that will be re-evaluated.

****Between the final data project and final exam, the higher grade between the two will given 30% of your final grade, and the lower grade will be given 25% of your final grade.**

Administrative Details

Submitting assignments: We will be using the turnitin.com system to submit, grade and review all homework assignments and the final data project. I will provide links to turnitin.com on the course website for you to submit your assignments. The midterm and final exam are multivariate-choice exams and will be administered in class on the days indicated above.

Students are allowed to discuss homework with one another. However, students cannot simply copy one another's work or turn in a joint project. The rule is that, after discussion, each student must do his or her own homework assignment independently. You may give each other advice about difficult problems and help find errors in one another's code, but in the end you must carry out the assignment yourself and not simply copy or use someone else's work. The turnitin system will automatically pick out assignments that are very similar, including those copied from the Internet so please do not copy or otherwise cheat on assignments. It is an easy way to get a zero in an assignment.

Late assignments: Late assignments will not be accepted. We may consider late assignments depending on circumstance; for instance a medical emergency backed by documentary evidence. The cut-off for a particular assignment is 11:59 PM on the day it is due.

Absence: It is highly recommended that you come to all class meetings. Each day's lecture will always build on the previous lectures and so you will certainly fall behind if you are absent for one class. More importantly, the R statistical package used for this class is likely to be new to most students. Without a good grasp of R programming techniques, you will almost certainly not be able to complete any of the assignments in this class.

Plagiarism: On homework assignments, you may work with *one* partner. If you do work with a partner, include your partner's name under yours in the heading. Working with a partner does not mean copying from each other – your submission *must* be your own work. In general, University-wide policies regarding plagiarism apply to this course with full force. They have been enforced in the past; if in doubt about something, ask an instructor. I will refer to the Dean of Students in the case cheating is suspected.

Consultation and Help

The TAs will lead the discussion sections and be available for consultation regarding R codes, data essays/home work assignments during office hours. Check the course website for their office locations and meeting times. Students are encouraged to use the discussion board on the class website to post questions and also respond to questions posted by other students. The TAs and I will not answer coding questions through email.

TAs

1. Shawn Patterson: shawnpattersonjr@gmail.com
 - Section A 3pm – 3:50pm
 - Section B 4pm – 4:50pm
 - Section C 5pm – 5:50pm
2. Estefania Casteñeda: ecp9@ucla.edu
 - Section D 9am – 9:50am
 - Section E 10am – 10:50am
 - Section F 4pm – 4:50 pm

Essentials and Expectations

- *Math*: Students should be comfortable with high school algebra II topics, especially functions, linear equations, matrices and graphs. Training in more advanced mathematical topics such as calculus and statistics is helpful but not required.
- *Writing*: A significant portion of the class will focus on effectively communicating the results of data analysis through written prose. Students should be comfortable with putting together assignments of 5-7 pages in length. The initial homework assignments will require fewer pages.
- *Computers*: A computer is necessary for almost everything we do in this class, except the midterm and final exams. Students are advised to bring a laptop to class during the first week of the quarter to follow along with the lectures. If needed, laptops can be borrowed from CLICC through this link: <http://www.library.ucla.edu/clicc/lending>. In subsequent weeks, much of the computer work will be done in sections so there will be no need for computers during the regular lecture times.
- *Software*: All statistical analyses will be conducted using the statistical package R, in conjunction with the RStudio integrated development environment (IDE). While the learning curve for R is steeper than for packages such as STATA or SPSS, it has myriad benefits that make it worthwhile: it is transparent, has a growing user base, and best of all, it is legally FREE. Students will be provided with sample code for all methods we cover. If you will be using your own computer, you can download and install the latest version of R for Mac [here](#) and for Windows [here](#). You should also download and install RStudio from [here](#) for both Mac and PC. Please be sure to have these programs installed before the second class meeting on Thursday, October 1. If you need help installing these programs, contact any of the TAs or ask the front desk folks at the Social Sciences Computing labs in the Public Affairs building.

The instructor will provide R-code examples sufficient enough to do the homework assignments. You may NOT email the instructor or TAs questions regarding your personal R-code for the homework assignments. We encourage all to post any questions about R-code onto the discussion board.

- Videos to help install R & RStudio:
 - PC: <https://www.youtube.com/watch?v=Tnhp65hxtZc>
 - Mac: <https://www.youtube.com/watch?v=eD07NznguA4>
- **Suggested Resources**
- UCLA ATS website < <http://www.ats.ucla.edu/stat/> >
- R Cookbook < <http://www.cookbook-r.com/> > (online book!)

Schedule of Topics and Assignments

Week 1	Oct 2	About PS6; Syllabus; Class information; installing R and R-Studio
	Oct 4	Introduction to R. Concepts. Measures of Central Tendency.
	Section	Installing and introduction to R.
<hr/>		
Week 2	Oct 9	Measures of spread. Histograms. Boxplots.
	Oct 11	Measures of spread. Types of distributions.
	Section	Histograms, boxplots, and subsetting data in R.
<hr/>		
Week 3	Oct 16	Hypothesis testing. Controlled comparisons: cross-tabulations, contingency tables.
	Oct 18	Bivariate relationships. Correlation. Scatterplots.
	Section	Subsetting data, log transformation, and tables.
		Data Essay #1 due October 20 at 11:59PM
<hr/>		
Week 4	Oct 23	Introduction to regressions analysis: bivariate regression, slope coefficients, R-squared:
	Oct 25	Regression analysis: Interpretation and assessment. Log variables.
	Section	Scatterplot and regression.
<hr/>		
Week 5	Oct 30	Midterm Review
	Nov 1	Midterm
		No section.

Data Essay #2 due November 3 at 11:59PM

Week 6

Nov 6 Data over time: Longitudinal data and line graphs**Nov 8** Introduction to multivariate linear regression; statistical control in observational studies..**Section** Line graphs; multivariate regression

Week 7

Nov 13 Multivariate regression: Interpretation and assessment. Dummy variables.**Nov 15** Central Limit Theorem. Statistical Inference.**Section** Multivariate regression.

Data Essay #3 due November 17 at 11:59PM

Week 8

Nov 20 Confidence Intervals and Standard Errors**Nov 22** No Class

No Section

Week 9

Nov 27 Different variable, Same Data: The consequences of variable choice on research design and data analysis**Nov 29** Binary Dependent Variables**Section** Multivariate regression and prep for Final Project

Week 10

Dec 4 Counterfactuals and Causal Inference.**Dec 6** In-class review for final**Section** Review and final project

Week 11

WEDNESDAY, December 13 from 8:00 AM to 11:00 AM

Finals Week

Dec 13

Location Didd 121

Research paper due on FRIDAY December 15, at 11:59PM