

Introduction to Data Analysis

Political Science 6

2015 Fall Quarter

Syllabus

| | |
|------------------------|---|
| <i>Meetings:</i> | Haines 220, Tu/Th 3:30PM - 4:45PM |
| <i>Course Website:</i> | https://moodle2.sscnet.ucla.edu/course/view/15F-POLSCI6-1 |
| <i>Instructor:</i> | Jesse Acevedo |
| <i>Email:</i> | jesseacevedo@ucla.edu |
| <i>Office Hours:</i> | Bunche 4276, Tu/Th 2:00PM – 3:00PM |

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 two data essays (i.e. homework assignments), a midterm, and a final consisting of two parts – a research paper and an in-class written 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 16 | 5 |
| Data Essay 2 | Oct 30 | 10 |
| Midterm Examination (In-class) | Nov 5 | 25 |
| Data Essay 3 | Nov 18 | 10 |
| Final data project: Research Paper | Dec 11 | 20 |
| Final Examination (In-class Exam) | Dec 8 | 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 first two data essays will be very structured; the third and final data project will be

open-ended. Note that 25 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.

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 written 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.

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.

TAs

1. **George Derpanopoulos** **gderpa@ucla.edu**
2. **Feng Yang** **yangfengnk@gmail.com**

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 8-10 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 (particularly October 1) to follow along with the lectures. If needed, laptops can be borrowed from CLICC through [this link](#). 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>

Recommended Text: This will be the main textbook for the class.

- Philip H. Pollock III (2012). *The Essentials of Political Analysis, 4th ed.*

Suggested Texts and Resources

- Moore D. S., McCabe G. P. 2006. *Introduction to the Practice of Statistics, 5th ed.*
- Illowsky B., Dean S. *Collaborative Statistics* <<http://cnx.org/content/col10522/latest>>
- Miller J. E. 2004. *The Chicago Guide to Writing about Numbers.*
- UCLA ATS website < <http://www.ats.ucla.edu/stat/> >
- R Cookbook < <http://www.cookbook-r.com/> > (online book!)

Schedule of Topics and Assignments

| | | |
|--------|---------------|--|
| Week 0 | Sep 24 | About PS6; Syllabus; Class information; installing R and R-Studio. |
| Week 1 | Sep 29 | Concepts and measurement; Measuring variables: levels of measurement; Describing variables: measures of central tendency – mean, median, mode. <i>Pollock: Introduction; Chapter 1; Chapter 2.</i> |
| | Oct 1 | Intro to R and RStudio, Loading packages (the “swirl” package), Loading and viewing data. |
| | | HW1 released (Due on October 16 at 11:59pm to turnitin.com) |
| Week 2 | Oct 6 | Describing variables: measures of spread – range, quartiles, IQR, percentiles. Describing patterns, what do you see? Histograms; boxplots. |
| | Oct 8 | Key terms: data, variables, parameters, and statistics. Shapes of distributions: Skewness Other measures of Spread: Variance, Standard Deviation |
| Week 3 | Oct 13 | Proposing explanations, framing and testing hypotheses: <i>Pollock Chapter 3.</i> |
| | Oct 15 | Controlled comparisons: cross-tabulations, contingency tables, interactions: <i>Pollock Chapter 5.</i> |
| | | HW2 released (Due on Oct 30 at 11:59pm to turnitin.com) |
| Week 4 | Oct 20 | Foundations of statistical inference: Random sampling, Central Limit Theorem and Normal distribution. Significance Tests. <i>Pollock Chapters 6-7</i> |
| | Oct 22 | Introduction to Bivariate Data: Correlation and Scatterplots |
| Week 5 | Oct 27 | Introduction to regressions analysis: bivariate regression, slope coefficients, R-squared: <i>Pollock Chapter 8, Pages 182-199.</i> |
| | Oct 29 | Interpreting and assessing regression. Log variables in regression analysis. |
| | | HW3 released (Due on November 18 at 11:59pm to turnitin.com) |
| Week 6 | Nov 3 | Midterm Review in class |
| | Nov 5 | Midterm in Haines 220, from 3:30 – 4:45pm |
| Week 7 | Nov 10 | Data over time: Longitudinal Data and line graphs. |

| | | |
|------------------------|-------------------|--|
| | Nov 12 | Different variable, Same Subject: The consequences of variable choice on research design and data analysis |
| | | Final project posted on Friday November 13. Due on Friday December 11 at 11:59PM to turnitin.com. |
| Week 8 | Nov 17 | Introduction to multiple linear regression; statistical control in observational studies. Interpretation and assessment. <i>Pollock Chapter 8: Page 199 – 211.</i> |
| | Nov 19 | Multiple Linear Regression continued: Dummy variables, standard errors, confidence intervals, model choice |
| Week 9 | Nov 24 | What can we do with our toolset?: Statistical inference and causal inference in PS6. Guidelines and expectations for final research paper |
| | Turkey Day | No class. |
| Week 10 | Dec 1 | Binary Dependent Variables |
| | Dec 3 | In-class review for final |
| | | Course review, final exam review. |
| | | Research paper due on Friday December 11, 2015 at 11:59PM to turnitin.com. |
| Week 11 Finals Week | Dec 8 | TUESDAY, December 8, from 11:30AM to 2:30PM Location TBA |
