University of California, Los Angeles
Department of Political Science

Introduction to Data Analysis

Political Science 6, 2014 Spring Quarter

Meetings:	1209B Bunche Hall, MW 10AM - 11:50 AM	
Course Website:	https://moodle2.sscnet.ucla.edu/course/view/14S-POLSCI6-1	
Course Email:	polisci6@polisci.ucla.edu	
Instructors:	Daniel Lim (weeks 1-5)	Joseph Asunka (weeks 6-10)
Office Hours:	Bunche 4250, MW 1:15-2:30 PM	Bunche 4250, MW 1:00-2:00 PM

This course introduces data analytic methods used by social scientists to make inferences about how the world works. This course has 3 objectives. First is 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). The second is 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. The final objective is to teach students how to present their data and results by effectively combining prose and graphical presentations (e.g. tables, histograms, scatterplots, boxplots). In pursuing these goals, we will explore political data from both developed and developing countries, on a variety of topics such as elections, economic development, and indicators of human welfare. In summary, this course will teach students how to think about and analyze data, and present their findings in a clear and compelling manner. .

Grading

Item	*(Due) Date*	*Percent of final grade*
HW 1	4/8	2
HW 2	4/15	3
HW 3	4/22	5
Midterm Exam	4/30	30
HW 4	5/16	10
HW 5	5/23	10
Final Exam – take home	6/6	20
Final Exam – in class	6/10	20

Homework assignments will involve producing statistical output and figures, analyzing them to find patterns, and writing essays about them. Initial assignments will be very structured; later ones will be open-ended. 30% of your grade is based on homework assignments, so you must take them just as seriously as the midterm and final exams. The best way to learn data analysis is to *do* data analysis; the homework assignments are designed to get you to do just that.

Administrative Details

Submitting assignments: We will be using TurnItIn to submit, grade and review all assignments. Log in through MyUCLA, register and get acquainted with the system.

Late assignments: Late assignments will lose 10% of their raw value for each day beyond the due date. The cut-off for a particular assignment is 11:59 PM on the day it is due.

Attendance: As this is a lecture course, attendance is not mandatory. However, it is *highly* recommended that you come to class. Each day's lecture will always build on the previous lectures, so you will almost certainly fall behind if you are absent for one class.

Sections: TAs will lead discussion sections and consult during those times, as well as during their own office hours, regarding R code, and data essays/homework assignments. As with lectures, sections are optional but *highly* recommended.

Essentials and Expectations

Math: Students should be comfortable with high school algebra II topics, especially functions, algebraic equations, matrices and graphs. Training in more advanced mathematical topics, such as calculus and statistics, is helpful but not required.

Writing: This class *requires* students to effectively communicate the results of data analysis through written prose of up to 8-10 pages in length.

Computers: A computer is necessary for most things in this class. Students should bring a laptop to class to follow along with the lectures. If needed, laptops can be borrowed from CLICC.

Software: Statistical analyses will be conducted using the statistical package R, in conjunction with the RStudio integrated development environment (IDE). The learning curve for R is steeper than for packages such as STATA or SPSS, but it has myriad benefits that make it worthwhile: it is transparent, has a growing user base, and best of all, is free. Students will be provided with sample code for all methods we cover. R may be downloaded here, and RStudio, here.

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 or TA.

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/ >

Schedule

3/31 **Introduction**
Admin, Intro to R, Loading and viewing data, Intro to election turnout dataset, Measures of central tendency

---

4/2 **Univariate I**
Histograms, Measures of spread, Boxplots

4/7 **Univariate II**
Population versus sample, skew, transforms

4/9 **Normal Distribution I**
Intro/motivation, visualization, PDF/CDF

4/14 **Normal Distribution II**
PDF/CDF plots, CLT

4/16 **Normal Distribution III**
Parameterization, Fitting to real data, 68-95-99 rule

4/21 **t-distribution**
Intro to t-distribution (and others), test of difference in means

---

4/23 **Bivariate I**
Intro to GDP dataset, contingency tables, scatter plots, line graphs

4/28 **Bivariate II / Midterm Review**
Covariance and correlation, in-class midterm review – section reviews TBD

4/30 **Midterm**
10AM, Bunche 1209B

5/5 **Bivariate III**
More on scatter plots. Dependent/independent variables, least squares line, residuals, slopes etc. (pocketbook voting data – RDI and Vote Share).

5/7 **Regression I**
Introduction to regression analysis (bivariate regression); slope coefficients, interpretation.

5/12 **Regression II**
More on bivariate regression: statistical significance, standard errors, p-values, substantive significance.

5/14 **Regression III**
Multiple regression; statistical control in observational studies.

5/19 **Regression IV**
More on multiple regression; hypotheses testing, writing results.

5/21 **Causation I**
Correlation versus causation, reverse causation.

5/28 **Causal designs**
Randomization, experimental design, causal inference.

6/2 **Regression wrap up**
Beyond linear regression: Brief introduction to non-linear models.

6/4 **Course / Final review**

6/10 **Final exam**
3-6PM, Bunche 1209B