University of California, Los Angeles

Department of Political Science

**Introduction to Data Analysis**

Political Science 6, 2013 Summer Session C

*Instructor*: Daniel Lim, [danielklim@ucla.edu](mailto:danielklim@ucla.edu)

*Meetings*: 1209B Bunche Hall, MW 10:45 AM-12:50 PM

*Office Hours*: Bunche 4250, MW 1:15-2:30 PM

*Course Website:* <https://moodle2.sscnet.ucla.edu/course/view/131C-POLSCI6-1>

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. Objective one 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 objective 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 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.

**Grading**

|  |  |  |
| --- | --- | --- |
| *Item* | *(Due) Date* | *Portion of Grade* |
| HW 1 | 8/13 | 10% |
| Quiz 1 | 8/19 | 15% |
| HW 2 | 8/20 | 15% |
| HW 3 | 8/27 | 15% |
| Quiz 2 | 9/4 | 15% |
| HW 4 | 9/5 | 15% |
| HW 5 | 9/13 | 15% |

**Prerequisites and Expectations**

* *Math*: Students must 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.
* *Computers*: A computer is necessary for most things we do in this class, except the quizzes. Students are advised to bring a laptop to class to follow along with the lecture notes. If needed, laptops can be borrowed from CLICC [<http://www.clicc.ucla.edu/tiki-index.php>].
* *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, is free. Students will be provided with sample code for all methods we cover.

**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/>>

**Administrative Details**

* *Submitting assignments*: We will be using the Turnitin system to submit, grade and review all assignments. When you log into MyUCLA, there will be an option for “Turnitin - Student” in the navigation panel. Please follow the link, 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.
* *Absence*: As this is a lecture course, attendance is not mandatory. However, it is highly recommended that you come to class, else you will almost certainly fall behind.

**Schedule**

|  |  |
| --- | --- |
| 8/5 | **Univariate I**  Intro to R, Loading and viewing data, population versus sample, Measures of central tendency, Histograms |
| 8/7 | **Univariate II**  Measures of spread, Boxplots, Subsetting data |
| 8/12 | **Univariate III**  Normal distribution I, Density, Density plots |
|  |  |
| 8/14 | **Univariate IV**  Normal distribution II, CLT, Random variables, t distribution, Hypotheses & difference of means |
| 8/19 | **Quiz 1**  **Bivariate I**  Data shapes (long versus wide), Transforming variables, Contingency tables, Scatterplots |
| 8/21 | **Bivariate II**  Line graphs, Correlation & covariance |
| 8/26 | **Linear Models I**  Bivariate regression, regression output |
| 8/28 | NO CLASS |
| 8/30 | Reschedule of 8/28 lecture  **Causality**  Association versus causation, statistical control, bias, experiments |
| 9/2 | Labor Day – reading assignment TBD |
| 9/4 | **Quiz 2**  **Linear Models II**  Multivariate regression, prediction |
| 9/9 | **Linear Models III**  Dummy variables, factors, interactions, graphing thereof |
| 9/11 | **Intro to GLM**  Typical use scenarios, comparison with LMs, examples |