

**The science of politics:  
An introduction to coding and statistical analysis in political science  
Harvard Pre-College Program  
Summer 2019**

**Sarah James**  
**sarahjames@g.harvard.edu**

**Course description**

Curious about how to interpret all of those (often contradictory) claims that you read in the media about government, politics, and policy? Understanding the science and statistics behind social science data analysis empowers you to analyze findings on your own. Social scientists use a wide array of statistical tools to test their claims and assumptions about the social world. In this course, you will be introduced to the core tools in this toolbox. You will learn how to calculate the certainty of your claims about hypotheses; code and use the open source statistical analysis program R; find and curate interesting data on the functioning of government and democracy; run and interpret regressions; and interpret and critique scientific data analyses. We conduct a limited exploration of as much of the “why and how” behind statistics as possible given that students will not have taken calculus.

In addition to learning basic statistics, we examine various measures and indices commonly used in political science. We will be working with datasets on a range of topics: the U.S. Congress, the developing world democracies and presidential elections, political participation, and US presidential election data. The topics covered in this course mirror those covered in quantitative methods courses for undergraduates and graduate students in the social sciences. You will also develop and answer your own political science research question using the analytical tools you gain in the course.

**Note that a laptop is required to participate in the course; however, if access to a laptop will be an issue, please contact me and we can work out a plan to provide technology for students who may need it.**

**Essential Questions**

- How should we assess claims made using data? In other words, what are best practices in data analysis and reporting that give us confidence the conclusions are valid?
- How should we interpret correlations?
- What is the difference between correlation and causation?
- How can we use quantitative data to test hypotheses about the functioning of government and democracy?

**Expectations**

This course mirrors the workload and structure of a college style course in quantitative methods in political science. I expect you to complete all assignments before the start of class and come prepared to both ask questions and engage in respectful, scholarly conversation. Working with statistical analysis programs, coding, and computers more generally can be both immensely frustrating and rewarding. I expect you to have patience with yourself and the program as you learn the ins-and-outs of using it. Remember that learning to code is like learning a new language. I also expect all assignments by the assigned deadline. Failure to meet deadlines will be reflected in your final course evaluation.

**Short term Assignments**

Syllabus—Quantitative Methods in Political Science  
Summer 2019

You will have three short problem sets due throughout the course. These problem sets will give you practice conducting the analysis that we discuss in class. You are also expected to be making nightly progress on your final research paper each night. Portions of the problem set may require you to conduct analysis on your respective data sets. If you are unable to arrive at an answer you are satisfied with for a problem on the problem sets, you must provide an explanation of what you tried and where you got stuck.

You will also have short project assignments on non-problem set nights. These assignments require incremental progress towards completing the long-term assignment. Each project task is described at the end of the syllabus.

### Long term Assignments

Throughout the course you will work on analyzing a dataset to answer a research question that you develop. You will have the option of using a dataset provided by the instructor or collecting the data on your own from an online source. We will reserve the last 30-60 minutes of each class and the entirety of day 9 as workshop time where you can work on your research project and get help from me and your peers. Details about the write up of your analysis will be provided on the first day of class. You will submit **your final research question and data source by midnight on Friday, July 26th**. The **final analysis will be due by 10am on the last day of the course**.

### Course Schedule

#### Part 1: The basics

#### **July 22: Introduction, Good Data Management & Intro to R**

- Download R studio
- Set the working directory and upload a dataset in R studio
- Identify the features of a good, analyzable dataset

#### **July 23: Mean, median, mode: Choosing the right statistic**

*Project task 1 due*

- Compute basic summary statistics in R
- Explain the advantages of different summary statistics (mean, median, mode, range)

*Readings:*

- Introduction to *Mastering Metrics*
- Skocpol, Theda, and Alexander Hertel-Fernandez. "The Koch network and Republican party extremism." *Perspectives on Politics* 14.3 (2016): 681-699.

#### **July 24: Causation vs. correlation**

*Problem Set 1 due*

- Describe the fundamental problem of causal inference
- Distinguish between causation and correlation
- Identify the pitfalls of relying on correlation as causation
- Filter data in R using dplyr

*Readings:*

- Read Chapter 1 of *Mastering Metrics*
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. "The colonial origins of comparative development: An empirical investigation." *American economic review* 91.5 (2001): 1369-1401.

#### **July 25: Frequentist statistics; omitted variable bias**

*Project task 2 due*

- Describe the assumptions behind frequentist statistics

Syllabus—Quantitative Methods in Political Science  
Summer 2019

- Identify sources of OVB and explain the impact on point estimates
- Describe expected value

*Readings*

- Acemoglu, Daron, Simon Johnson, and James A. Robinson. "A response to Albouy's 'A reexamination based on improved settler mortality data'." *Unpublished working paper* (2005).

**July 26: Hypothesis testing & data sources**

*Problem Set 2 due*

*Right Wing Authoritarianism Scale*

- Explain the significance of a difference in means test
- Calculate the t-test and explain its meaning
- Create scatterplots to examine trends in data
- Explore sources of open data, available for analyzing

*Readings*

- Mcdermott, R., Tingley, D., & Hatemi, P. (2014). Assortative Mating on Ideology Could Operate Through Olfactory Cues. *American Journal Of Political Science*, 58(4), 997-1005.
- Hetherington, Marc, and Elizabeth Suhay. "Authoritarianism, threat, and Americans' support for the war on terror." *American Journal of Political Science* 55.3 (2011): 546-560.

**Part 2: Regression**

**July 29: Introduction to bi-variate regression**

*Project task 3 due*

*DW-Nominate*

- Interpret a bi-variate regression
- Interpret a p-value and explain its meaning
- Run a bi-variate regression in R
- Reading: bi-variate regression

*Readings*

- Chapter 2 of *Mastering Metrics*
- Carnes, Nicholas. "Does the numerical underrepresentation of the working class in Congress matter?." *Legislative Studies Quarterly* 37.1 (2012): 5-34.

**July 30: Introduction to multi-variate regression**

*Problem Set 3 due*

*Freedom House Scores*

- Interpret a multivariate regression
- Run a multivariate regression in R
- Explain how to interpret binary dependent variables

*Readings*

- Armstrong, David A. "Stability and change in the Freedom House political rights and civil liberties measures." *Journal of Peace Research* 48.5 (2011): 653-662.
- Whitten-Woodring, Jenifer. "Watchdog or Lapdog? Media Freedom, Regime Type, and Government Respect for Human Rights." *International Studies Quarterly* 53 (2009), 595-625

**July 31: Regression diagnostics**

*Project task 4 due*

*Racial Resentment index*

- Use various statistical tests to check the validity of your regression
- Identify ways to adjust your regression in response to diagnostic results
- Properly present charts and graphs and reference them in scholarly papers

*Readings*

- Henderson, Michael, and D. Sunshine Hillygus. "The dynamics of health care opinion, 2008–2010: Partisanship, self-interest, and racial resentment." *Journal of Health Politics, Policy and Law* 36.6 (2011): 945-960.

**Part 3: Beyond regression & wrap up**

**August 1: Workshop day**

*Project task 5 due*

- Workshop presentation and paper

**August 2: Presentations**

*Final analysis and project due*

- Present findings to your peers
- Give feedback to your peers on their research

**Project tasks**

**Task 1**

Part 1

- Look through the datasets available for use for the final projects, or look for your own.
- Identify 2 that you would be interested in working with.
- In a word document, identify the two datasets you are interested in working with, and brainstorm 2-3 questions that you could answer using the data from each dataset.
- If you find your own dataset, explain whether it is well organized for analysis. If it is not, what changes do you need to make?

Part 2

- Using your state or local government's website, find a publicly available dataset
- Describe the structure of the dataset, and explain whether the structure of the dataset is well organized for analysis. If it is not, explain how you would restructure it to make it better

**Task 2**

- Choose the dataset that you plan to work with for the final project, and explain what question you plan to answer with your dataset.
- Provide summary statistics for the key variables in your dataset.

**Task 3**

- Identify potential sources of OVB in your dataset
- Identify the model that you plan to run to answer your research question.
- Try to run your model, then identify any errors, questions or challenges you are facing in doing so.

**Task 4**

- Provide the results from the model that you ran to answer your research question.
- Interpret the coefficients
- Identify any questions or concerns that you have about running your model.

**Task 5**

- Create a to-do list of everything you need to accomplish in order to turn your final project in on time.

- Identify which tasks you plan to get done during class on Day 9, and mark any tasks for which you need teacher or peer assistance.

### **Academic Integrity**

Plagiarism is the theft of someone else's ideas and work. It is the incorporation of facts, ideas, or specific language that are not common knowledge, are taken from another source, and are not properly cited.

Whether you copy verbatim or simply rephrase the ideas of another without properly acknowledging the source, the theft is the same. A computer program written as part of your academic work is, like a paper, expected to be your original work and subject to the same standards of representation. In the preparation of work submitted to meet course, program, or school requirements—whether a draft or a final version of a paper, project, take-home exam, computer program, placement exams, application essay, oral presentation, or other work—you must take great care to distinguish your own ideas and language from information derived from sources. Sources include published and unpublished primary and secondary materials, the Internet, and information and opinions of other people.

You are expected to follow the standards of proper citation and to avoid plagiarism. Please consult the [Harvard Guide to Using Sources](#), prepared by the Harvard College Writing Program, for a helpful introduction to all matters related to source use: identifying and evaluating secondary sources, incorporating them into your work, documenting them correctly, and avoiding plagiarism.

### **Accessibility services**

*The Harvard Pre-College Program and I are committed to providing an accessible academic and residential community.*

The accessibility services office offers a variety of accommodations and services to students with documented disabilities, permanent and temporary injuries, and chronic conditions. If you are a student with a disability, we engage you in an interactive process to provide you an equal opportunity to participate in, contribute to, and benefit from our academic and residential programs.

The manager of accessibility services works with you, your instructor(s), and staff on an individualized, case-by-case basis to provide appropriate services to ensure you have a rich and rewarding academic and campus experience.

You can find additional information at <https://www.summer.harvard.edu/resources-policies/accessibility-services>.