2024

Quantitative Analysis in Sociology: OER Syllabus

Matthew McKeever
Haverford College, mmckeever@haverford.edu

Quantitative Analysis in Sociology – OER Syllabus
Matthew McKeever

The objective of this OER syllabus is to provide a no-cost open access course that introduces students to the use of quantitative data in sociology. The material covered in the syllabus examines statistical methods used to analyze data. Statistics textbooks tend to be expensive, yet examine topics that have not changed much in the past few decades. Consequently, this course relies on the discussion of statistical topics from online sources. There are several sources listed for each topic, so that multiple ways of presenting the material are available to students. The syllabus consists of 10 units, so that it can be adapted to either a quarter or semester schedule. During a full semester, the course includes extended discussion of assignments, introduction of additional statistical topics, and covers measurement theory as it applies to social science data.

Supplemental to covering statistical analysis, the course also asks students to use statistical software and publicly available social science data to conduct analysis on topics of interest. In the course at Haverford, class time is spent teaching students how to use Stata, with the option to supplement the instruction with videos from the Stata YouTube channel. Stata is not required, however, and any free software program that allows for the same types of analyses (such as R) is acceptable.

Over the course of the semester there are exams to test that concepts are clear to everyone, and a set of short papers that allow them to practice writing about data analysis. The course builds up to a final research piece paper that requires students to identify a research question, a dataset that can be used to answer it, and an analysis of the data. As such, the course reproduces the process of writing a social scientific research article.

Sociology 215
Quantitative Methods
(First Implementation: Spring, 2024)

This course is an introduction to the methods used to analyze quantitative data in sociology. The course objectives include learning to think about the uses of different types of quantitative data in sociology, to calculate and analyze simple and multivariate statistics using statistical software, and to present your findings in a standard research paper format. The idea behind these is that the best way to gain an understanding of how researchers use such data to make sociological arguments is to practice doing it yourself. The course is written with sociology students in mind, but these methods apply equally well to other disciplines in the social sciences.

The course is structured to reproduce a standard process of developing a quantitative research project in the social sciences. In each section of the course statistical methods are reviewed, then used to analyze social science data. For this reason, the class includes both discussion and lab activities, with weekly assignments that require reporting on the data analysis completed each week.

The statistical methods that we will be reviewing are very commonly taught across a wide variety of disciplines, and have for the most part changed little over the past half-century. The math behind the calculation of the statistics we'll be examining is the same in a textbook from fifty years ago as it is today. For that reason, there's not any one recent textbook that's necessary for achieving the course objectives, and it's definitely not worth paying a lot of money for one. There are many statistical textbooks available for free across the internet, and the course builds on a set of them. I've created an outline that draws on six different sources that covers the statistical concepts you'll be expected to understand to complete the final project:

- Carnegie-Mellon Open Learning Initiative [OLI], www.oli.cmu.edu, course on Probability and Statistics
- *OnlineStatBook*, http://onlinestatbook.com/
- Holbrook, Thomas. An Introduction to Political and Social Data Analysis using R. https://bookdown.org/tomholbrook12/bookdown-demo/
- *Probability and Statistics Ebook*, http://wiki.stat.ucla.edu/socr/index.php/EBook (you might have to get to this through the web archive. If so try this link: https://web.archive.org/web/20160312104059/http://wiki.stat.ucla.edu/socr/index.php/EBook#Chapter_I:_Introduction_to_Statistics)
- Openstax *Introductory Statistics*, https://openstax.org/details/books/introductory-statistics
- *OpenIntro Statistics*, by D. Diez, M. Cetinkaya-Rundel, and C. Barr, https://www.openintro.org/book/os/

Although many are listed, you might find that only one or two are really necessary for achieving the course objectives. Most of these are directly accessible. For the Carnegie-Mellon course, you can either pay to access the course, or access it for free by registering as an "Independent Learner" through their web page: https://oli.cmu.edu/independent-learner-courses/ . For some weeks, there are additional supplementary readings, all of which should be available for free. In addition, you can consult a social statistics textbook. There are plenty of additional examples of ones online, and you can always pick one up on the cheap from various used book sellers.

While the statistics used to describe data have changed little over the past century, the approach to quantitative analysis – how data and statistical methods are discussed and utilized – has changed a great deal. For that reason, the primary focus of the class time is to discuss what you find as you use these methods. There will be opportunities to write analysis nearly every week so that you may become familiar with the conventions of writing quantitative social data analysis.

Finally, the course will require you to learn a statistical program to analyze data. The course materials will reference one in particular, Stata. There are many standard programs in use in both academic and non-academic research. The most common programs include R, SAS, Stata, and SPSS, but there are many more programs with a smaller user base that are also sufficient for conducting introductory analysis. It is possible to analyze data using any of these, and so there is no particular one required for this course. Stata is referenced in the course because it is relatively easy to learn, and yet includes just about any statistical procedure researchers use in the social sciences. There is a large user base for Stata, and so there's a great deal of help available to use it online. This includes notes from other statistics classes, online help forms, and even a YouTube channel from the company that makes the program; https://www.youtube.com/user/StataCorp
.
Another common program, R, is freely available through at: https://education.rstudio.com/. The user base for R is slightly larger than that for Stata, and so consequently there is a great deal of support online for this program as well. In the past students have found this program a bit more challenging to learn at the beginning, though the long-term benefits of using it might be worth the initial investment in time if the professional field you're interested in is more likely to use this program over Stata.

Data: To complete the course requirements, you will need to use survey that will allow you to empirically examine a research question of your choosing. To start off the course, we will use a subset of the General Social Survey to showcase how one analyzes data. More on these data can be found at the GSS website: https://gss.norc.org You can watch a short introductory video here: https://youtu.be/I94Do1Lwm_o. As the course proceeds, you will be encouraged to download a dataset that will allow you to complete a final project of interest to you. More on how this can be accomplished will be discussed as the course proceeds.

Course Topics and Readings

Topic 1	Statistics and the logic of data analysis.
Readings	*OLI Probability and Statistics*: Unit 1, Module 3, page 9: "The Big Picture." Holbrook: Chapter 1. *OnlineStatBook*: Chapter I (Introduction), sections A-D [aka sections 2-5] *Probability and Statistics Ebook* : Chapter 2, section 2.2 Openstax *Introductory Statistics*: Chapter 1, section 1.1
Topic 2	Measurement, distribution of data, measures of central tendency, and variation.
Readings	*OLI Probability and Statistics*: Unit 2, pages 10-37

	Holbrook: Chapters 3, 5, 6 *OnlineStatBook*: Chapter I (Introduction), sections F, H, I [aka sections 7, 9-11], Chapter II (Graphing Distributions), Chapter III (Summarizing Distributions), sections 1-9, 12-15. *Probability and Statistics Ebook* : Chapter 3. Openstax *Introductory Statistics*: Chapter 1, sections 1.2 & 1.3, Chapter 2 *OpenIntro Statistics* : Chapter 1, section 2-2.4, Chapter 2, section 2.18
Topic 3	Bivariate distributions: crosstabulation, measures of association, and scatterplots
Reading	*OLI Probability and Statistics*: Unit 2, pages 38-51 Holbrook: Chapter 13. *OnlineStatBook* : Chapter IV (Describing bivariate data) section 2. Supplemental reading on Phi and risk ratios: https://web.pdx.edu/~newsomj/cdaclass/ho_phi.pdf
Topic 4	Sampling, inference, and hypothesis testing
Reading	*OLI Probability and Statistics*: Unit 3, Module 6, pages 67-71, Module 7, pages 84-87, Unit 4, Module 12, pages 164-172, Unit 5, Model 13, pp 176-178. Holbrook : Chapters 8, 9 *OnlineStatBook* : Chapter 9, sections 1-6, Chapter 11 *Probability and Statistics Ebook* : Chapter 6, sections 1-2, Chapter 8, section 1. Openstax *Introductory Statistics*: Chapter 7, sections 1-3
Topic 5	Significant bivariate relationships
Reading	*OLI Probability and Statistics*: Unit 5, Module 14, pages 215-224, Module 15, pages 225-226, Module 17, pp 255-259. Holbrook : Chapter 10 *OnlineStatBook* : Chapter 12, sections 1-4, Chapter 17, *Probability and Statistics Ebook* : Chapter 9, section 2 Openstax *Introductory Statistics*: Chapter 9, sections 1 & 2, Chapter 11 *OpenIntro Statistics* : Chapter 6, sections 6.3, 6.4, Chapter 7, section 7.3
Topic 6	Regression
Reading	*OLI Probability and Statistics*: Unit 2, pages 52-59. Holbrook: Chapter 15 *OnlineStatBook* : Chapter 14, sections 1-5 *Probability and Statistics Ebook* : Chapter 10, sections 1-2 Openstax *Introductory Statistics*: Chapter 12, sections 1-3 *OpenIntro Statistics* : Chapter 8, sections 8.1-8.3

Topic 7	Regression and Inference
Reading	*OnlineStatBook* : Chapter 14, sections 6-8 *Probability and Statistics Ebook* : Chapter 3 Openstax *Introductory Statistics*: Chapter 12, sections 4-5. *OpenIntro Statistics* : Chapter 8, section 8.4
Topic 8	Multivariate regression
Reading	Holbrook: Chapters 16-18 *OnlineStatBook* : Chapter 14, sections 9-11 *Probability and Statistics Ebook* : Chapter 10, section 4, Chapter 15, section 8 on 'Fisher's F-distribution' *OpenIntro Statistics* : Chapter 9, section 9.1, 9.3, 9.4 Some supplementary readings on dummy variables : Taboga, Marco (2021). "Dummy variable", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/fundamentals-of-statistics/dummy-variable Berman H.B., "Dummy Variables in Regression." Available at: https://stattrek.com/multiple-regression/dummy-variables. https://www.statology.org/dummy-variables-regression/ https://blogs.ubc.ca/datawithstata/home-page/regression/ordinary-least-square/ On F-tests: https://www3.nd.edu/~rwilliam/stats1/x94.pdf
Topic 9	Logistic regression
Reading	*OpenIntro Statistics* : Chapter 9, section 9.5 Supplementary readings https://stats.oarc.ucla.edu/stata/dae/logistic-regression/
Topic 10	Additional statistical techniques
Reading	There are many additional types of regression analysis that we do not cover in this course. Most deal with categorical dependent variables. Examples of many of these regressions, with examples, can be found at : https://stats.oarc.ucla.edu/other/dae/.