POL 502: INTERMEDIATE STATISTICS FOR PUBLIC POLICY

Spring 2019
Mondays 5.00 – 8.00 pm
Room: SBS N703
Stony Brook University

Instructor: Vittorio Mérola
Office: SBS N721
Office Hours: Mondays 2 - 4 pm, and by appointment
Email: vittorio.merola@stonybrook.edu

Course Description

This is the second course in the statistical sequence for the Masters of Public Policy. In this course, you will be introduced to regression models, both linear and non-linear. These are the workhorse models of political science and policy research. This course will also provide the foundation for a more general understanding of quantitative methodology. As such, you will learn how to **draw inferences** from samples to populations, how to formally **test hypotheses** using data, and how to **evaluate data-based claims** in observational research. This is both a theoretical and practical course, with **hands-on experience in analyzing data**, which will involve learning how to apply the techniques we discuss in class using a popular and powerful statistical software program, R.

At the successful completion of the course, you should be able to **critically evaluate the data analysis in most social scientific research** you come across; and **seek out and analyze data** to address research questions that interest you. The goal of the course is to teach you to understand and to confidently apply a variety of statistical methods that are essential for political science and public policy research.

Prerequisites

While much of this course is devoted to learning methods of statistical analysis, it is not necessary for you to have an extensive mathematical background in order to gain a thorough understanding of the material.

However, there are a couple of factors you should keep in mind when deciding if this course is the right fit for you:

- To be successful in the course, you should expect to devote considerable (though not unreasonable) time to study the material each week. Learning to use statistics is not easy, and everyone struggles at first. But the rewards, both in terms of your knowledge and analytical ability, as well as your job prospects, are definitely worth it!
- Learning how to use R can be a frustrating process. You should be willing to put in the time and effort to use it, since using R is an intergral part of the course. That said, I am here to help you learn and will, of course, provide extensive information on how to perform data analysis in R.

Texts and Material

Required Texts

The following free textbooks are recommended. I will assign chapters from these books for the various weeks of this course, although this is to serve as background material, since all the required concepts and applications will be covered in lecture.

- David M. Diez, Christopher D. Barr, and Mine Cetinkaya-Rundel. *OpenIntro Statistics*. CreateSpace, 3rd edition.

  A PDF of the book (henceforth OI) is available for free at `https://www.openintro.org/stat/os3.php`.

- Gareth James, Daniela Witten, TRevor Hastie and Robert Tibshirani. . *An Introduction to Statistical Learning, with Applications in R*. Springer.

  A PDF of the book (henceforth ISL) is available for free at `http://www-bcf.usc.edu/~gareth/ISL/`.

If you want a more advanced free textbook, which is more mathematically based and coveres the relevant topics at greater length, I recommend *Advanced Data Analysis from an Elementary Point of View*, by Cosma Rohilla Shalizi, which is available for free at `http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf`. I will suggest some readings from this book (henceforth ADA) for the more advanced topics in this class, but this is purely for those of you who are curious and want a more thorough understanding of those topics. As I said, everything you need to know for this class will be covered in lecture.

Additional Resources

`http://stattrek.com` – Stat Trek is a helpful website for statistics. It has a help page on many of the statistics topics we'll cover and you should consult it as a secondary source on statistics topics that are giving you trouble.

`https://stats.stackexchange.com/` – Stack Exchange is another helpful website for statistics. It is a site where you can ask questions, and find answers to commonly asked questions. The answers are not always easy-to-understand, but it is a useful resource as your knowledge of statistics advances.

Computer Software

Quantitative social science research requires the use of computers. Throughout the course, I will assume that you have a personal (or family) computer that you can use to complete all assignments. **If you do not have access to a personal computer, please email me ASAP.**

While there are many different statistical software packages available, we will be using a program called R, which is a free, open source programming language and environment for statistical computing. It can be downloaded at http://www.r-project.org/. To assist you in using R, I also recommend that you install RStudio, which is a free, open source interface to R that makes working with R significantly easier, particularly on Windows. It can be download it for various platforms at http://www.rstudio.org/. I will provide instructions on how to install both programs in week 1.

R Help

It takes a while to properly learn how to use R. Thankfully, there are an almost infinite amount of resources online to help you learn R. In this class, we will spend a significant amount of time going through code and learning how to use R during class time. Importantly, we will go through the necessary codes that you will need to complete the homeworks and to estimate the tests and models that will be covered during lecture. However, if you still need additional help understanding R, or if you would like to deepen your knowledge of R beyond what we cover in class, the following free resources should be helpful:

- `http://www.statmethods.net/index.html` – Quick-R is a help website for problems you may encounter when using R.
- `https://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf` – A good introduction to R programming.
- `https://cran.r-project.org/doc/manuals/R-intro.pdf` – A slightly more in-depth introduction to R programming.
- `https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf` – A free book teaching you statistics through R
- `https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf` – A very comprehensive, but accessible, overview of R.
- `https://leanpub.com/rprogramming` – A fairly comprehensive book on the nuts and bolts of R.
- `http://www.burns-stat.com/pages/Tutor/R_inferno.pdf` – A more advanced R programming book.
- `https://r4ds.had.co.nz/index.html/` – Another more advanced R programming book.
- `https://resources.rstudio.com/` – Useful videos and tutorials, with an RStudio focus.
- `http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/` – Great introduction to plotting in R using ggplot2.

- `https://cran.r-project.org/doc/contrib/Short-refcard.pdf` – A quick and useful reference card for commonly used commands.
- `http://scs.math.yorku.ca/index.php/R:_Getting_started_with_R` – A useful site for learning R and statistics more generally.
- `http://scs.math.yorku.ca/index.php/R:_Getting_started_with_R` – A bunch of additional resources.
- `https://stackoverflow.com/questions/tagged/r` – A good place to ask questions and find answers to questions.

Assignments and Grading

Assessment for this course will be based on 5 homeworks and 1 final project. The final course grade will be determined based on the following breakdown:

Homework assignments: 5 @ 10% each = 50%
Final paper: 1 @ 35% = 35%
Attendance: 15 @ 1% = 15%

Homework

Homework assignments are meant to assist you in developing the problem-solving, analytic, and computer skills necessary to perform modern social scientific research. Assignments will give you the opportunity to engage deeply with the course material and provide you with hands-on experience in working with real-world data in R. The five homeworks will comprise 50% of your final grade. The schedule below contains information on when the assignments are due. The homeworks should be written up and handed in, or emailed, by the beginning of the class on the date when they are due. They will be posted on blackboard at least 1 week before they are due, along with the data and codebook required for the analysis.

You are allowed, and even encouraged, to work together on the homeworks. It is good practice to first try to develop answers on your own, and then meet in a group setting to discuss potential difficulties. However, the final write-up of the answers and the computer code which you hand in should be written by you. DO NOT simply copy computer code or answers from your classmates.

Final Project

You will also write a final paper, about 15-20 pages long (double-spaced), which will perform statistical analyses on some data of your choosing. It will be worth 35% of your final grade. This can be the analysis on some original data that you have collected, some existing data that you have obtained, or even previously published data which you would like to replicate or extend. You are free to use statistical methods beyond those learned in class, but the

majority of the analysis should utilize concepts we developed during the semester. All the R code used to produce the results in the paper should be included as an appendix (which does not count for the page length) at the end of the paper.

Please come see me if you have any questions about this, or if you need help finding data or deciding on the right project. The final paper will be due 1 week after the last class of the semester, on **May 13**.

Attendance Policy

We will meet once a week during the semester. You can expect me to be prepared, give the lecture and answer questions. When you come to class, please also be prepared. I will require attendance. There are 16 lectures in the course, of which you can miss 1 during the semester without it affecting your grade - unless you have a valid reason for missing class. Attendance comprises 15% of you final grade. The classroom is a great place to exchange ideas, meet your classmates, and ask questions. Regular attendance is vital since lectures are tightly linked to the homeworks, and will cover material that is not presented in the textbook. In particular, we will spend a significant amount of time in lecture going through examples in R, in order for you to learn both the theory behind the statistics and the computer code necessary to implement it. If you do not attend regularly, it will be difficult to pass the class.

Tips for Success

1. Come to class each week and ask questions. Statistics takes time to learn, and requires a lot of repetition and persistence, not least through question and answer.

2. Get an early start on homework assignments. Using R will involve some trial-and-error and having ample time to seek help from the instructor is crucial.

3. Don't be a stranger! I am here to help and am happy to answer your questions. Please come by my office hours, even if your questions seem small (which they never are).

4. Work together on the homeworks. Talking things through with other people is the best way to truly learn difficult material.

5. Use all the resources available. If I can't help, and the book isn't useful, look things up online. I can guarantee you that you can find the answer online to pretty much any question you will have about the material, regardless if it is about R code or statistics.

Class Schedule

The class schedule below is tentative, and is subject to change depending on how fast we get through the material. If the schedule does change, an updated syllabus will be posted to blackboard.

Week	Dates	Topics	Assignments
1	1/28	**Introduction**	
2	2/4	**Probability Theory** Read: OI #1.6-1.8, #2, #3.1-3.2, #3.4	
3	2/11	**Statistical Inference** Read: OI #4.1-4.5, #5, #6.1-6.4	
4	2/18	**Bivariate Regression** Read: OI #7, & ISL #3.1	HW1 Due
5	2/25	**Multiple Regression I** Read: OI #8.1, & ISL #3.2, #3.2.1	
6	3/4	**Multiple Regression II** Read: OI 8.2, & ISL #3.2.2, #3.3.1, #3.3.2	HW2 Due
7	3/11	**Assumptions & Diagnostics** Read: OI #8.3, & ISL #3.3.3	
8	3/18	**Spring Break - NO CLASS**	
9	3/25	**NO CLASS**	
10	4/1	**Hypothesis Testing** Read: ISL #3.4, #3.6	HW3 Due
11	4/8	**Advanced Regression Topics** Read: ISL #7.1-7.6 (ADA #4, #7)	
12	4/15	**Likelihood Functions and GLMs** Read: ISL #4.1-4.2 (ADA #11.2-11.4)	HW4 Due
13	4/22	**Logistic Regression** Read: OI #8.4, & ISL #4.3	
14	4/29	**Ordered and Multinomial Models** Read: ISL #4.4	
15	5/6	**Advanced Topics - LAST CLASS** Read: ISL #5.2 (ADA #6)	HW5 Due
16	5/13	**Final Paper**	Due by midnight