PS 230: Introduction to Political Research

University of Illinois at Urbana-Champaign
Spring 2017
January 1, 2017 to May 4, 2017

331 Armory
Monday and Wednesday 9:00–9:50am

Instructors

Dr. Jake Bowers
jwbowers@illinois.edu
Office: 432 David Kinley Hall
Office Hours: Monday and Wednesday 10:00–10:50am or by appointment

Wei Zhong
Office: 309 David Kinley Hall
Office Hours: Friday 3:00–6:00pm or by appointment.

Sections:
Friday 11:00 – 11:50 am G23 in Foreign language building
Friday 2:00 – 2:50 pm G7 in Foreign language building

Course Description

This class is an introduction to applied statistics as practiced in political science. It is computing intensive, and, as such, will enable students to execute basic quantitative analyses of social science data using the linear model with statistical inference arising from resampling and permutation based techniques as applied in the R statistical computing language. By the end of the course, a successful student will be able to explore social science data online, download it, analyze it, and write about how the analyses bear on focused social science or policy questions.

Goals and Expectations

More than anything I assume a willingness to engage with mathematics, data analysis, computer programming, and the practice of social science thinking and writing. I also assume you've taken at least one class in algebra at the level taught in most high schools in the United States and have used a personal computer to read and type email and other documents and have some experience with the Internet.

I also assume that you will read the syllabus and that you keep up to date on changes in the syllabus which will be announced in class). You should not expect a response to emails that ask a question already answered in the syllabus.

Over a number of semesters, the many people involved in designing this course have allowed for some flexibility. This means that it is likely that the syllabus will change throughout the term. Make sure you have the syllabus with the latest date stamp. I will also announce syllabus changes via emails sent from the Moodle.

If you have any special needs (for example any disability that you' d like us to know about) please contact me during the first two weeks of class. I am happy to help and to work with the folks from `http://www.disability.illinois.edu/`

Course Website

You can find this syllabus, as well as course readings and assignments at: `https://learn.illinois.edu/course/view.php?id=21774/`

Use the following URL to access an online version of the RStudio Integrated Development Environment (IDE): `https://rstudio.atlas.illinois.edu/`

Course Requirements:

In Class Work

The class itself will involve work in groups at your computers nearly every class meeting. This is not a lecture class but a hands-on learning experience. I may assign new groups throughout the semester if it looks like some people are not learning as much as they should. At the beginning of each class, you will download a worksheets with problems that will require you to work as a group to use the R statistical computing language. The problems will be designed first to introduce you to the idea of scientific computing as practiced in the social sciences and then to the basics of social science data analysis and frequentist statistical inference. You will work on the worksheets together during the class-time (in groups of about 3 people), but each individual is responsible for uploading their own answers. We will grade one problem from each classwork selected at random.

Participation

Quality participation does not mean "talking a lot." It includes attending section; thinking and caring about the material; and expressing your thoughts respectfully and succinctly and thoughtfully. Participation, in this class, will mostly refer to your active involvement in your sections, but the quality of your general involvement in lectures, emails, and office hours will also be taken into account. If you are late for class, expect your participation grade to suffer.

Final Report:

Each of you will write a final paper no longer than about 10 pages. This paper is an opportunity for you to use the ideas from this class to pursue some data analysis on a topic that interests you. I will have several

assignments and class sessions oriented around your paper to (1) give you practice with the techniques under discussion and (2) push your paper along so that the quality of papers turned in at the end is high. I will also require you to turn in two partial drafts of your paper—thus ensuring that you do not have to scramble at the end of the term to complete your paper and that you have some input on your work. I will be working through the format of the final paper as the class proceeds. Roughly speaking, in the final paper, I will expect you to put together what you've learned in class with your own interests to execute a simple bit of statistical data analysis (including fitting a linear regression model, and graphing and interpreting the results). For example, you will produce a regression table and be able to explain what p-values and confidence intervals mean as well as the substantive meaning of the regression coefficients.

Grading and Course Polices

Grades are comprised of the following:

- 40% In-Class Work and Attendance This part of the grade consists of 70% in-class work grade and 30% attendance. I will drop the lowest 3 of the daily worksheet grades. The worksheets will require you to do open-ended data analysis to arrive at the correct answer although the answers will tend to require very little writing. If you answer the randomly chosen problem correctly you will receive an A on that worksheet (100%). If you answer incorrectly you will receive a B on that worksheet (86.99%). If you do not answer the question chosen for grading, you will receive 0%. Obviously, if you do not attend the class that day, you will receive a zero for your worksheet grade. Attendance will be a simple percentage of the number of class sessions you attended. In-class work happens in-class. It may not be turned in late or made-up at a later date without official excuses: For example, if you are hospitalized in the middle of the term, but the Dean thinks that you should not drop the course, I will work with you, your doctors and the relevant Dean to enable you to complete the course.
- 30% Final Reports Grades on the final reports will be based on the clarity of your writing and thinking and the correctness of your data analysis. You may turn in reports late, but you will lose $\frac{1}{3}$ letter grade for each day that you are late (e.g. an "A" assignment would become a "A-" assignment after 1 day, a "B+" assignment after two days, . . . , a "C" assignment after 6 days). The Final proposals are a part of the Final Report, and, as such the Final Report Grade will be `max(Final Report*.80+Final Proposal*.20,Final Report)`. The final proposal grade itself will be calculated in the same way as `max(Final Proposal*.80+Draft Proposal*.20, Final Proposal)`. As you can see, I aim to reward improvement.
- 20% Quizzes. During the semester, I will administer at least three surprise quizzes. In general, these will be harder than the class assignments but should not be too difficult if you keep up with the reading and make an effort during class.
- 10% Participation. This part of the grade has reflected both attendance and quality participation in class and also useful conversations during office hours or email as well as my sense about whether you did the readings and have had a constructive and engaged attitude toward the course during the main twice-weekly class meetings.

Incomplete Work Assignments not turned in will be counted as zero in the calculation of the final grade.

Computers in class You should bring a laptop to class or plan to work with someone who has a laptop. We will be using a cloud-based version of the RStudio software. If you would like to install an offline version, you can find the R statistical software and RStudio IDE online with instructions for local installation. I can help troubleshoot initial installation issues, but recommend that most students use the online version.

Books

Required:

- Kaplan, D. (2012). Statistical Modeling A Fresh Approach. Daniel Kaplan, Macalester College, St. Paul, MN, second edition [Called "ISM" for the rest of the syllabus. The first few chapters are available online for free at http://www.mosaic-web.org/go/Statistical Modeling/index.html.]

Recommended:

- Gonick, L. and Smith, W. (1993). The cartoon guide to statistics. Harper Perennial New York, NY [Nice coverage of hypothesis testing and confidence intervals as well as other topics at a very accessible level.]
- Verzani, J. (2005). Using R for Introductory Statistics. Chapman & Hall/CRC [Another nice textbook combining statistics with R. (see http://wiener.math.csi.cuny.edu/UsingR for more materials related to this book.)]
- Becker, H. S. (1986). Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article. University of Chicago Press [A wonderful book on social science writing.]
- Abelson, R. (1995). Statistics as Principled Argument. Lawrence Erlbaum, New York [Provides some very useful frameworks for how one might use statistics within the context of doing scholarly work.]

Academic Honesty:

Neither the University nor I tolerate cheating or plagiarism. If caught, you will face punishment under the guidelines provided by the University of Illinois. I expect you to be familiar with and understand the university's policies on academic honesty for this course. Please consult the Student Code for more information.

`http://www.admin.illinois.edu/policy/code/article1_part4_1-401.html`

We will discuss specific information about your written work in class in more detail, but if you are unsure of how to properly cite material, please ask for clarification. If you are having difficulty with writing or would like more information or assistance, the UIUC Writers' Workshop is a great resource. Please see their website for more information. `http://www.cws.illinois.edu/workshop/`

Community Standards

Political discussions can sometimes grow contentious. All students and the instructor must be respectful of others in the classroom. If you ever feel that the classroom environment is discouraging your participation or problematic in any way, please contact me immediately.

Office Hours:

I can answer many questions you may have through email, but more detailed questions about course policies, lecture material and readings, or about political science in general may be better addressed in person. I am available to meet immediately after class, by appointment, or in my office (432 David Kinley Hall) from

10:00 to 10:50 am on Monday and Wednesday. I ask that you email me if you plan on stopping by for office hours. I may be elsewhere on campus or have other meetings scheduled with other students or collaborators if you do not let me know that you plan to come by in advance.

Computing

In this class, we will be using the R statistical language (`http://www.r-project.org`) and the RStudio integrated development environment for R (`http://rstudio.org`). This means that we will be learning some computer programming skills. We will be typing sequences of commands in the R language in a text editor (`http://en.wikipedia.org/wiki/Text_editor`) and then asking the R interpreter to execute these commands. We will not be pointing and clicking to execute statistical analyses.

Computing is an essential part of modern statistical data analysis—both for producing persuasive information from data and for conveying that information to decision makers. So we will pay attention to computing, with special emphasis on understanding what is going on behind the scenes.

The final reports must be turned in on the class Moodle either as pdf, postscript, or html. Documents in Microsoft Word format (or Wordperfect, or Pages, or OpenOffice) will not be accepted.

Emergency Response

In the unlikely event of an emergency, campus policy is that students, faculty and staff should "run, hide, fight." That is, persons in danger should first seek to distance themselves from danger, conceal themselves indoors if unable to flee, and as a last resort, defend themselves. More information about the university's emergency response can be located here `http://police.illinois.edu/emergency/response/`.

Course Schedule

This schedule is preliminary and subject to change. If you miss a class make sure you contact me or one of your colleagues to find out about changes in the lesson plans or assignments.

Week 1

Wednesday, January 18: Introduction and course overview

No readings

Week 2

Monday, January 23: What do we mean when we say "data"?

- **Task** Make sure that your RStudio online account is working. Contact the instructors. if there are issues.

- **Read** ISM Chap. 1 and 2

Wednesday, January 25: Why do we want to talk about variation?

- **Read** ISM Chap. 3

Week 3

Monday, January 30: Why do we want to talk about variation?

- **Read** ISM Chap. 3

Wednesday, February 1: Why do we care about models?

- **Read** ISM Chap. 6 (ISM Chap 4 as Background)

Week 4

Monday, February 6: Why do we care about models?

- **Read** ISM Chap. 6 (ISM Chap 4 as Background)

Wednesday, February 8: What is a linear model? How are linear models useful?

- **Read** ISM Chap. 7

Week 5

Monday, February 13: What is a linear model? How are linear models useful?

- **Read** ISM Chap. 7

Wednesday, February 15: How can we fit linear models to data?

- **Read** ISM Chap. 8

Week 6

Monday, February 20: How can we fit linear models to data?

- **Read** ISM Chap. 8

Wednesday, February 22: Quiz 1

Week 7

Monday, February 27: Model Fit and R^2

- **Read** ISM Chap. 9

Wednesday, March 1: Statistical Adjustment: Holding Constant

- **Read** ISM Chap. 10

Week 8

Monday, March 6: Holding Constant by Subsetting or Stratifying

- **Read** ISM Chap. 9

Wednesday, March 8: Holding Constant by Residualization, Covariance Adjustment, "Controlling For"

- **Read** ISM Chap. 10

Week 9

Monday, March 13: When is holding constant meaningful?

- **Read** Read Berk 2004, Chapter 6.5 and 7.2

Wednesday, March 15: Quiz 2

Week 10

March 20: NO CLASS SPRING BREAK

March 22: NO CLASS SPRING BREAK

Week 11

Monday, March 27: How can we talk about uncertainty about (or confidence in) our models? What is a "sampling distribution"? Confidence intervals.?

- **Read** Chap. 12 (and ISM Chap 5 as background)

Wednesday, March 29: Report Proposal Workshop

Task Come to class prepared to work on your report proposal (i.e. your plan for writing your report).

Friday, March 31: Report Proposals Due by 5pm

The report proposal is a proposal: it tells the reader what question you have and what dataset and variables you **plan** to use. The point of this assignment is to get you thinking seriously about your final paper and doing the hard work of finding data, grappling with codebooks, and thinking about how outcome variables, explanatory variables, and control variables relate to your questions. Specifics about the form of this proposal (and draft report and final report itself) will be handed out in class as the term progresses.

Week 12

Monday, April 3: Sampling Distributions and model uncertainty

- **Read** Chap. 12 (and ISM Chap 5 as background)

Wednesday, April 5: What is a hypothesis test?

- **Read** Chap. 13,14,15

Week 13

Monday, April 10: Why test a hypothesis about a parameter in a model?

- **Read** Chap. 13,14,15

Wednesday, April 12: Report Draft Workshop

- **Task** Come to class ready to work on the draft of the first half of your report.

Friday, April 14: Report Drafts Due by 5pm

Week 14

Monday, April 17: Can math can make our lives easier?

- **Read** Freedman, Pisani and Purves 2007, Chap 16âĂŞ18. The Law of Large Numbers and the Central Limit Theorem.

Wednesday, April 19: Do we always have to resample/permute?

- **Read** Freedman, Pisani and Purves 2007, Chap 16âĂŞ18. The Law of Large Numbers and the Central Limit Theorem

Week 15

Monday, April 24: Hypotheses about whole models? Evaluating fit. More Predictive Plotting and Checking

- **Read** ISM Chap. 14

Wednesday, April 26: Quiz 3

Week 16

Monday, May 1: Putting it all together: Report Writing

- **Task** Come to class ready to work on your final reports.

Wednesday, May 3: Putting it all together: Report Writing

- **Task** Come to class ready to work on your final reports.

Friday, May 5: Final Report Due by 5pm

- **Task** Review and options for undergraduate research, internships in data science, and other coursework.