Political Science 230
Introduction to Statistics for Political Scientists

Jake Bowers
jwbowers@illinois.edu
Moodle: `https://moodle.atlas.uiuc.edu/course/view.php?name=09SpPS230`

Spring 2009

General Information

This class is an introduction to applied statistics as practiced in political science.

Location and Time The whole class meets in 319 Gregory Hall for lectures on Mondays and Wednesdays from 11:00 AM to 11:50 AM.

Sections meet in G23 Foreign Languages Bldg on Friday mornings.

Office Hours Jake's office hours are 12–1 PM Mondays and Wednesdays in 495 Lincoln Hall. I will have three 20 minute meetings available for each office hour: You can choose 12:00–12:20, 12:20–12:40, or 12:40–1:00. Please make an appointment if you want to come to office hours.

Aya's office hours are TBA in 407 Lincoln Hall.

Goals and Expectations

More than anything we assume a willingness to engage with mathematics, data analysis, computer programming, and the practice of social science thinking and writing. We also assume you've taken at least one class in algebra at the level taught in most high schools in the United States.

Grades We'll calculate your grade for the course this way: Research paper (40%), Homeworks (40%), Participation (20%).

Paper Each of you will write a final paper due on May 12th. 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.

We will have several assignments 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.

Paper Grades will be based on a draft and a final paper. You will also receive some input in the form of grades (or unsatisfactory/satisfactory/excellent) on the small paper assignments that you will complete along the way.

Part of the grade for the paper will be based on the quality of the draft that you present. The proportion of the paper grade based on the draft will vary, depending on when you turn in your draft. People who turn in their drafts earlier, with less time to work on their drafts, will have less of their paper grade determined by their draft. People turning in drafts later will have more. The precise breakdown will be: 15% of the paper grade for those who turn in their drafts in group 1, 20% of the paper grade for groups 2 or 3, 25% for groups 4 or 5.

If your final paper grade is higher than your draft grade, your paper grade will reflect only the final paper grade.

If your final paper grade is lower than your draft grade, your paper grade will be the average of the two grades, weighted equally.

Homeworks We will have ≈4 problem sets due throughout the term. You will be assigned into groups of ≈5 students at random within your Friday section.

We will probably to assign groups using the following method. For example, here we assign 20 students to 4 groups:

```
> load("ISM.Rdata")
> students <- 1:20
> groups <- gl(4, 5)
> set.seed(1234)
> finalgroups <- sample(groups)
> rbind(students, finalgroups)[, order(finalgroups)]

            [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
students       1    7    8   12   15    6    9   10   11    16     2     3
finalgroups    1    1    1    1    1    2    2    2    2     2     3     3
            [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20]
students        5    14    20     4    13    17    18    19
finalgroups     3     3     3     4     4     4     4     4

> table(finalgroups)

finalgroups
1 2 3 4
5 5 5 5
```

Notice that it is random (in that if we did it again we would get a different result), but that we are guaranteed 4 groups of 5 people in each group:

```
> finalgroups <- sample(groups)
> rbind(students, finalgroups)[, order(finalgroups)]

            [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
students       3    4    5   11   15    1    2    7    9    13     6     8
finalgroups    1    1    1    1    1    2    2    2    2     2     3     3
            [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20]
students       17    18    20    10    12    14    16    19
finalgroups     3     3     3     4     4     4     4     4

> table(finalgroups)

finalgroups
1 2 3 4
5 5 5 5
```

We could also roll a 4-sided die:

```
> finalgroups.dice <- resample(groups)
> rbind(students, finalgroups.dice)[, order(finalgroups.dice)]
```

```
                 [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
students            9   11   15   19    3    5    8   12   17     1     2     4
finalgroups.dice    1    1    1    1    2    2    2    2    2     3     3     3
                 [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20]
students             6     7    13    14    16    10    18    20
finalgroups.dice     3     3     3     3     3     4     4     4
```

But even though each of the 4 groups has equal probability of being selected for a given person, a given set of 20 rolls is not guaranteed to provide exactly 5 1s, 5 2s, etc... Although there are always exactly 20 people distributed among groups.

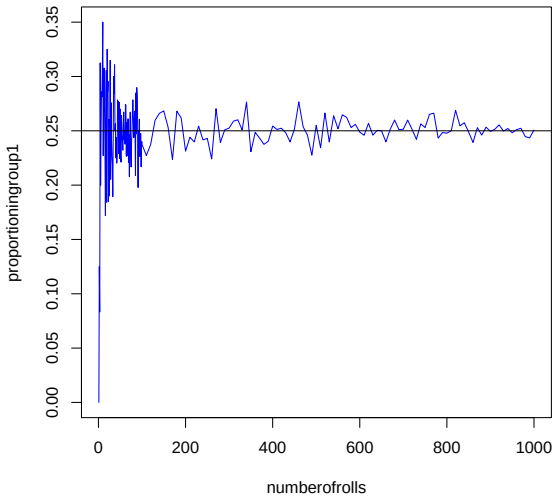
```
> table(finalgroups.dice)

finalgroups.dice
1 2 3 4
4 5 8 3

> sum(table(finalgroups.dice))

[1] 20
```

The dice option only works if we have a very very large class requiring many dice rolls:

```
> finalgroups.bigN.dice <- resample(gl(4, 5))
> table(finalgroups.bigN.dice)/20

finalgroups.bigN.dice
   1    2    3    4
0.40 0.20 0.25 0.15

> finalgroups.bigN.dice <- resample(gl(4, 500))
> table(finalgroups.bigN.dice)/2000

finalgroups.bigN.dice
     1      2      3      4
0.2485 0.2515 0.2585 0.2415
```

Here is a plot which shows how the distribution of students in groups corresponds to the number of times the dice is rolled (by the end of the class you'll guess that this code is unnecessarily inefficient):

```
> numberofrolls <- c(seq(1, 100), seq(100, 1000, 10))
> proportioningroup1 <- vector(length = length(numberofrolls))
> for (i in 1:length(numberofrolls)) {
+     numrolls <- numberofrolls[i]
+     assignedgroups <- resample(gl(4, numrolls))
+     if (any(levels(assignedgroups) == "1")) {
+         proportioningroup1[i] <- sum(assignedgroups == "1")/length(assignedgroups)
+     }
+     else {
+         proportioningroup1[i] <- NA
+     }
+ }
> plot(numberofrolls, proportioningroup1, type = "l", col = "blue")
> abline(h = 1/4)
```



Bonus question: Why did I need that `if(){}else{}` piece of the code?

Homework Grades The overall homework grade will be calculated as an average of the grades of homeworks by your groups (everyone in the group receives the same grade), including a grade gauging your performance as a group member given to you by yourself and the other members in your group (different for each member of each group).

Participation Quality participation does not mean "talking a lot." It includes turning in assignments on time; thinking and caring about the material; and expressing your thoughts respectfully and succinctly in class. Participation, in this class, will mostly refer to your active involvement in your sections, but your general involvement in lectures will also be taken into account.

Books

Required: Kaplan, D. (2008). *Introduction to Statistical Modeling.* `http://www.lulu.com`

Recommended: Gonick, L. and Smith, W. (1993). *The cartoon guide to statistics.* HarperPerennial 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. We will be grading the final papers under the assumption that you write the way Becker advises us to write.

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.

Computing

In this class, we will be using the R statistical language.

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.

All data analytic work will be turned in with an appendix that we can run (not cut and paste, but submit as a batch job) to replicate your analyses.

All written work will be turned in either as hard-copy, pdf, postscript, or html. Documents in Microsoft Word format (or Wordperfect, or Pages, or OpenOffice) will not be accepted.

Schedule

Note: 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.

January 21 — A Taste of the Course

Assignments

For section: Read ISM § 1.4. Bring your laptops if you have them. Important to get R installed.

For class: Start thinking about cool data and variables for your research paper.

January 26 — What does it mean to use statistics to answer political questions and Scientific Computing using R

Reading

ISM Chap. 1

January 28 — What do we mean when we say "data"?

Reading

ISM Chap. 2

Assignments

Distribute paper assignment 1

February 2 — Why do we want to talk about variation?

Reading

ISM Chap. 3

Assignments

Problem set 1 due.

February 4 — Why do we want to talk about variation?

Reading

ISM Chap. 3

Assignments

February 9 — Why do we care about models?

Reading

ISM Chap. 4

Assignments

Paper assignment 1 due. Hand out paper assignment 2 on describing variation and relationships.

February 11 — Why do we care about models?

Reading

ISM Chap. 4

Assignments

February 16 — What is a linear model? How are linear models useful?

Reading

ISM Chap. 5

Assignments

Paper assignment 2 due. Hand out homework 2.

February 18 — What is a linear model? How are linear models useful?

Reading

ISM Chap. 5

Assignments

February 23 — How can we fit linear models to data?

Reading

ISM Chap. 6

Assignments

Homework 2 due.

February 25 — How can we fit linear models to data?

Reading

ISM Chap. 6

Assignments

March 2 — Correlation ...

Reading

ISM Chap. 7

Assignments

March 4 — ... vs. Causation

Reading

ISM Chap. 8

Assignments

March 9 — Why "hold constant" and what does this mean anyway?

Reading

ISM Chap. 8

Assignments

Hand out paper assignment 3 on linear models and holding constant.

March 11 — The "holding constant" problem.

Reading

ISM Chap. 8

Assignments

March 16 — Getting ready for holding constant in linear models and for hypothesis testing: Model vectors

Reading

ISM Chap. 9

Assignments

Paper assignment 3 due.

March 18 — Model vectors

Reading

ISM Chap. 10

Assignments

March 23 & 25 — Spring Break

March 30 — Multiple Model Vectors

Reading

ISM Chap. 11

Assignments

Hand out homework 3.

April 1 — Multiple Model Vectors

Reading

ISM Chap. 11

Assignments

April 6 — Why model randomness?

Reading

ISM Chap. 12

Assignments

Homework 3 due. Hand out paper assignment 4.

April 8 — Why model randomness?

Reading

ISM Chap. 12

Assignments

April 13 — Random model vectors

Reading

ISM Chap. 13

Assignments

Paper assignment 4 due. Hand out homework 4.

April 15 — Random model vectors

Reading

ISM Chap. 13

Assignments

April 20 — How to talk about uncertainty for models? What is a "sampling distribution"?

Reading

ISM Chap. 14

Assignments

Homework 4 due.

April 22 — Uncertainty for models.

Reading

ISM Chap. 14

Assignments

First group of paper drafts due.

Section: Draft presentations TBA.

April 27 — Why test a hypothesis?

Reading

ISM Chap. 15

Assignments

Second group of paper drafts due.

Section: Draft presentations TBA.

April 29 — Why test a hypothesis?

Reading

ISM Chap. 15

Assignments

Third group of paper drafts due.

Section: Draft presentations TBA.

May 4 — Hypotheses about whole models? Evaluating fit.

Reading

ISM Chap. 16

Assignments

Fourth group of paper drafts due.

Section: Draft presentations TBA.

May 6 — Hypotheses about whole models? Evaluating fit.

Reading

ISM Chap. 16

Assignments

Fifth group of paper drafts due.

Section: Draft presentations TBA.

May 12 — Final Papers Due