Course description

The main goal of the course is to develop (i) understanding, (ii) ability to critically evaluate, and (iii) ability to confidently apply statistical analyses of the type covered in the course in order to answer substantive questions in political science.

The course will cover the classical linear regression (including assumptions, properties of estimators, violations of assumptions and solutions, tests, interpretation, extensions, and the like.) Toward the end of the course, we will also introduce in brief maximum likelihood and models of qualitative dependent variable.

The course should give you tools to assess what is an appropriate estimation technique by which to analyze your data, and, no less important, what are the pitfalls of using particular techniques versus others.

Books and reading materials

The following books are on reserve and available for purchase at the COOP:

We also put on reserve the following books:
They might come in handy in the matrix algebra section of the course.
Different people find different texts intuitive and helpful for different topics. I list below a few statistics/econometrics textbooks. I will occasionally refer to them. Please take the time to browse through them and find the ones helpful to you. These books are on reserve:


And these are a few additional ones:

*Substantive readings/applications/additional readings*. I weaved into the course plan substantive readings which are excellent examples of the topics learned. These readings are marked with *. A good example of an application goes a long way in demonstrating how a method is used and what its advantages are. We will discuss these readings in class. Please make sure to come prepared.

Our main textbook for the course is Greene’s. However, on some of the earlier weeks we will use other texts. Also, for every topic, I list below Greene some alternative readings from other textbooks should you prefer to consult with them. It is important that you read before the lecture.

We will have a mailing list for the class. Please make use of it to ask and answer each other’s questions. We all learn from each other’s questions.

**Assignments**

**Weekly problem set.** Problem sets will be handed on recitation and will be due the following recitation at the beginning of the session. They will include empirical and theoretical questions, depending on the topic. You may work in groups but do the write-up on your own.

The data we will use for most problem sets is the Comparative Study of Electoral Systems. The CSES is a terrific data set which allows for investigation of a variety of questions. It is a multi-country dataset including information both at the micro level about individuals and at the macro level about political systems. We will ask you to
focus on different parts of it depending on the week. The data are available at: www.cses.org. Please go ahead and acquaint yourself with these data.

Midterm exam. This will be a take-home exam, to take place on Wednesday, April 1st. It will be a 48-hour exam or more. Please plan accordingly.

Research paper. Research paper in which students will conduct original research. More details will be provided in class. Papers are due on Monday, May 18 at 4PM. Heads up: on Thu/Fri., April 16/17, as part of the weekly assignment, we will ask you to demonstrate initial progress on the research paper.

Draft of research paper. A rough draft is due on May 1\textsuperscript{nd}. Please hand in two copies (to us and to an assigned peer).

Peer commentary. Each student will be assigned to a peer and will provide commentary on the draft. The commentary should be constructive and aim at improvement of the work read. Please hand in two copies of the commentary (to us and to the assigned peer). The commentary is due on May 6 in class.

Grading. Weekly problem set - 20%, midterm exam – 30%, paper draft + peer commentary 15%, final paper - 35%.

Course plan

Wednesday, February 4  
**Introduction**

Monday, February 9  
**Probability and Statistical Inference - Review**

bias, consistency, efficiency

Greene, C1-C5
Gujarati, A1-A4, A6-A8
S+W, 2.1, 2.2, 2.5, 3.1, 3.2, 3.3

For recitation:


**Linear Regression - Bivariate Model**

Wednesday, February 11  
Least Squares assumptions

Tuesday, February 17  
model fit
Monday schedule

properties: finite sample, asymptotic

Gujarati, Ch. 2, 3
S+W, Ch. 4
Begin reading Achen, Sage monograph

Wednesday, February 18  **Linear Regression – Multivariate Model**
Gauss-Markov
assumptions and problems (no solutions yet)
model fit
properties

Gujarati, 4.1-4.3, 7.1-7.8
S+W, Ch. 5.4, 5.5, 6.2-6.6
Complete Achen, Sage monograph.

Monday, February 23 – Wednesday, February 25  **Review of Matrix Algebra**
Vectors, matrices, addition, multiplication, identity, inversion, rank, dependence and independence, partition.

Greene, Appendix A
Johnston, Ch. 4
Simon and Blume, Ch. 6, 7, 8 (partition)
Strang, Ch. 1, 2

Monday, March 2  **Linear Regression Model in Matrix Form**
Greene, Ch. 2, 3.1-3.2, 3.5, 4.4, 4.8, 4.9

Wednesday, March 4 – Monday, March 9  **Linear Regression**
confidence intervals, hypothesis testing
restrictions on coefficients
transformations, non-linearity

Greene, Ch. 4.6-4.7, 5.1-5.3, 5.6, 6.3
Gujarati, Ch. 8
S+W, 5.1-5.2, 7.1-7.2 (homoskedasticity only), 8.2

Wednesday, March 11- Monday, March 16  **Linear Regression**
dummy variables, interaction terms
predictions
interpretation

Greene, 5.6, 6.1-6.2
Gujarati, 9.1-9.6
S+W, 5.3, 8.3
Wednesday, March 18  
**Linear Regression**

Plots, graphs, and common mistakes


Monday, March 23  
No class, spring break

Wednesday, March 25  
No class, spring break

Monday, March 30  
catch-up and review

Wednesday, April 1  
midterm take-home exam. (This is a 48-hour exam or more. Please plan accordingly.)

Monday, April 6  
**Problems, Violations of Assumptions, Solutions**

outliers  
missing data  
collinearity


Greene, 4.8.1, 4.8.2  
S+W, 6.7  
Gujarati, 10.1-10.5, 10.7-10.9
Wednesday, April 8, Monday, April 13, Wednesday, April 15

**More Problems**

- heteroskedasticity
- correlated disturbances

Greene, 8.4-8.7
Gujarati, 11.1-11.7, 12.1-12.4, 12.6

- measurement error
- omitted-variable bias
- Instrumental variable

Greene, 12.1-12.5
Gujarati, 7.7-7.8
S+W, 6.1, 7.5, Ch. 12

Monday, April 20

- No class, Patriots Day

Wednesday, April 22, Monday, April 27

**Endogeneity, Simultaneous Equations**

Greene, 12.1-12.5 (continued)
S+W, 6.1, 7.5, Ch. 12 (continued)


Wednesday, April 29, Monday, May 4

**Maximum Likelihood**

- dichotomous dependent variable
- Logit, Probit

King, Ch. 4, Ch. 5.1
Long, 2.6, 4.1

Wednesday, May 6

**Logit and Probit**

- quantities of interest

King, 5.2
Long, 3.1-3.5

Monday, May 11
Wednesday, May 13

**Multinomial Choice Models**

MNL, CL, IIA

Maddala, 2.10-2.12
Long, 6.1-6.3, 6.7-6.8