Following on POS 514 and 516, or another course on research questions/design, this course offers an introduction to the most common data analysis techniques. In POS 514 and 516, you received a solid introduction to the basics of designing a research project and you worked on the important first steps, developing a research question and examining the advantages and disadvantages of alternative approaches/research designs appropriate for answering the relevant question. Given that background, we will take the next step and examine in detail one of the several analysis techniques/tools we have already introduced, statistical/quantitative analysis. We will build up from the beginning, examining univariate statistics, bivariate and multiple regression analysis, and techniques of statistical inference. The models we consider will be increasingly sophisticated and powerful. As time permits, we will also overview some more advanced statistical methods.

Because of its popularity, flexibility and importance for more advanced work, the heart of the course will highlight regression analysis, considering the basics, building models of increasingly complex specification and analyzing assumptions upon which the regression model (ordinary least squares) (and subsequent alternatives to the regression framework) are based.

At the outset, please remember that social statistics need to be considered in a broader perspective. Quantitative analysis is only one, albeit an important one, of several tools available to you as you focus on research. A clearly stated research question accompanied by an appropriate research design set up the statistics, which should be viewed as the tool for answering your research question (as well as an end in themselves). The final stage in the research process of course is interpreting the statistics. Because we are social scientists, we will put a heavy emphasis on interpretation, critiquing published articles using statistics and continuing the 516 tradition of developing your critical thinking skills as a social scientist. Thus, calculation
methods and numbers are an important part of the story, but we need to additionally place them in a broader perspective.

Therefore, also note that, though the purpose of this class is to insure you come away with a fair degree of mathematical/statistical knowledge, no prior statistical work is assumed. Based on past behavior, it is understood that many students approach their intro stats courses with considerable trepidation. So although there is much material to be covered, we will slow down or speed up according to student needs. Variation in enthusiasm and ability will be respected. At the other extreme, if students have had appropriate statistical background, we will move more quickly.

Course Goals

It follows from the above that by the end of the course, you will be better able to:

- Appreciate the role of quantitative techniques as a tool to systematically answer social science research questions.
- Understand and interpret commonly used statistical techniques/tools
- Critique with a sharper eye social science research as you read journals and books
- Become familiar with the logic or basic mathematics underlying some important and commonly used statistical techniques
- Work comfortably with a commonly used social science computer package (Stata)
- understand the basics and appreciate the additional possibilities available should you need more sophisticated analysis tools
- have Fun!!

**Remember:
- An approach highlighting curiosity—rather than trepidation—befits a social/political scientist. Stop and think. What am I really interested in? What do I really want to know?
- especially as a large class, you are not in this alone!
- You are first and foremost a social/political scientist using statistics as a tool to answer important research questions. You can always seek input on technical aspects of the work. You don’t have to know it all!

Class Structure

Each week, class will have a similar structure. This structure was created with the hopes of furthering your understanding of quantitative methods generally and also their application to a variety of research topics and your ability to apply them.

The structure of each class will be as follows:

- Introduction via a brief exercise,
- a discussion/lecture on the topic/statistical technique for the week
- a group/class exercise applying the statistical technique
- a conclusion particularly reiterating strengths and weaknesses of the technique
- an assignment due the following week reviewing the material
Course Requirements

- Weekly Assignments: (combination of problem sets, annotated STATA work, reaction papers); 20% reaction papers and Stata work due at Monday classes; problem sets due by Thursdays at 5 if you need more time.
- 2 Midterms: 15% each; due dates 3/15 (Friday before spring break) and 4/22, negotiable if necessary
- Final exam or final paper: 20% (approx 10-12 pages)
- Class Participation, exercises, attendance: 20%. (attendance required unless excused)
- Additional extra assignments (bringing readings or additional info about stats to class): 10%

Books
There are many good statistics and econometrics books out. I am asking you to purchase (at as low a price as possible) Studenmund and Pollock (a user friendly intro to STATA). Chapters from Moore and McCabe are on Blackboard as are several Sage (little Green book) classics. Studenmund is concise, provides uniformity and stands as a good reference. Moore and McCabe provide more social science oriented examples and perhaps a little more explanation. The Sage books are short, interpretive and written for social scientists. We will be working at various points from all. No need to spend a lot of money here; used or somewhat older editions are fine, and much is on Blackboard.


Sage Publications (a series of short “green” books covering a wide variety of statistical methods (http://www.sagepub.com; Quantitative Applications in the Social Sciences). We will be using several:


*Expect additional reading especially for Part III of the class.
Useful Websites:
http://www.stata.com/ - the Stata website
http://polmeth.wustl.edu/ - The Society for Political Methodology
http://wps.aw.com/awstudenmund/useecon_5/- the textbook’s website (includes datasets)
http://bcst.whfreeman.com/ips6e/default.asp?s=&n=&i=&v=&o=&ns=0&uid=0&rau=0
Moore and McCabe text website
The Interuniversity Consortium for Political and Social Research (ICPSR) website
http://www.icpsr.umich.edu/icpsrweb/landing.jsp
Notes on Course Policies:

**Students with Disabilities:** If you need any class accommodations due to a disability, please utilize University resources as needed, and please let the professor know well in advance of any assignments requiring accommodation(s). It is a function of the University to provide such accommodations as needed.

**Plagiarism and Academic Honesty**

The strength of the university depends on academic and personal integrity. In this course, you must be honest and truthful. Plagiarism is the use of someone else’s work, words, or ideas as if they were your own without giving the original author credit by citing him or her. If you have any questions about plagiarism, please contact me before submitting assignments for grading. Plagiarism violations will result in disciplinary action.

Additionally, it goes without saying that students are expected to meet the broader standards of academic honesty expected of students at a major university.

See link on University policy: [http://www.albany.edu/eas/104/penalty.htm](http://www.albany.edu/eas/104/penalty.htm)

See also the UAlbany Library “Plagiarism 101” Tutorial: [http://library/albany.edu/usered/tut.html](http://library/albany.edu/usered/tut.html)
Weekly Topics and Assignments
(More so than other classes, subject to adjustment based on student needs and interests—
but we will stay on track!)

Part I: Getting Started

Week 1, 1/28: Overall Introduction and Introduction to Univariate Statistics and
STATA
Where does quantitative methods fit into the overall conceptualization of a research
process? Why do people think the level of measurement of variables is so important
Why bother with univariate statistics? What are the advantages and disadvantages of
alternative univariate statistics?

Week 2, 2/4: Math Refresher and Practice with Univariate Statistics (graphical and
numerical representations)
* Moore, sections 1.1, 1.2
*Pollock, ch. 1, 2
*Wilkinson and Picket, The Spirit Level (2009), ch. 1-3, 5
*Reaction paper to the Wilkinson reading: In one to two pages, summarize and critique
the argument. What were they trying to do, how did they go about doing it, and what do
you think of their work? Explain.

Part II: The Regression Model and Other Basics

Week 3. 2/11 Intro to Bivariate Regression/Ordinary Least Squares
What is the purpose of the linear regression model? What procedures are used to
estimate model coefficients and why? What statistics are important in the interpretation
of output from a statistical package, and what does each tell you? What assumptions
need to be satisfied for the results of regression to be viable?
*Moore, ch 2
*Pollock, ch. 8
*Studenmund, ch. 1
*Lewis-Beck ch. 1 and 2
*Univariate Statistics Assignment due

Week 4, 2/18 Bivariate Regression Practice and intro to Multiple regression
What does it mean to add independent variables to a regression model, controlling for
additional variables? What are the similarities and differences between the bivariate and
multiple regression models, and why are they important?
Similarities and differences between bivariate and multiple regression?
*Studenmund, ch. 2, 3
*article
*Bivariate Regression Assignment due

Week 5, 2/25: hypothesis testing: basic concepts, logic, steps in Process
What is so important about statistical inference? How does the process work, and what is the rationale for the procedures? What are the advantages and disadvantages of significance tests? (This week, we will also begin an examination of cross-tabulations.)

*Moore, ch. 1.3, 5.2, 6.1, 6.2
*Pollock ch. 4 and 5
*Regression assignment #2 due

Week 6, 3-4: Practice and Extensions: significance tests and Cross-tabulations
Moore, ch. 7
*Studenmund, ch. 4,5
*Pollock, ch. 3, 6, 7
*statistical inference assignment due

3/11 Review/summary of Part II
*mid term due 3/15

3/18 spring break

Part III: Regression Assumptions

Week 7, 3/25: choosing Independent Variables and Multicolinearity
Why is it so important to understand the assumptions upon which regression is based? What is choosing independent variables so important? What is multicollinearity, how do you diagnose it, and what are some ways you can take it into account in your statistical models?

*Studenmund, ch. 6, 8
*Lewis-Beck, relevant sections

Week 8: 4/1: Functional Form, Dummy Variables and Interactions
What do you do if you think the relationships between your variables are nonlinear? Interactive? How do you handle nominal or ordinal independent variables in your work? Why is it so important to expand your models to include these possibilities, and, technically, how do you incorporate these additions to the models into your analysis?
*Studenmund, ch. 7
*Lewis-Beck, relevant sections
*Pollock, ch. 9

Week 10, 4/15: Dichotomous Dependent Variables: Logistic Regression
What do you do if your dependent variable isn’t based on interval data? Why do we need to learn yet another statistical technique, logit analysis? How does it work, and how is it similar or different than least squares regression?
*Studenmund ch. 13
*Pollock, ch. 10
Part IV: Culmination; you made it!—read/critique articles of your choice, intro to more advanced methods, begin work on your final project.

Week 11, 4/22  article critiques
*second midterm due
Week 12, 4/29: overview of some “advanced” methods; time series, factor analysis, simultaneous equation models etc.
*Studenmund, ch. 9, 10
*on time series debates in international relations:

- “Back to the Future: Modeling Time Dependence in Binary Data” by David B. Carter and Curtis S. Signorino
- “Time is Not A Theoretical Variable” by Nathaniel Beck

Week 13, 5/6: wrap-up

**Recommended** for those interested:


Bashaw, W. L. *Mathematics For Statistics*


Steps in Research Process

1. Project Conceptualization

   a. Project justification: Why is this an important topic?

   b. The Research Question: A 'why' question asking why there is variation in your dependent variable or phenomenon.

   c. Concept Definition: Define the major concept of your project (usually this means defining your dependent variable). Be clear about your definition. Think about the extent of agreement there would be on your definition. Think about possible alternative definitions and explain why yours is best.

   d. Hypotheses: What independent variables explain variation in your dependent variable? Be clear about the direction and type of hypothesis (positive/negative relationship, linear/curvilinear), and give three reasons why each independent variable should be related to your dependent variable.

   e. Rival hypotheses: This will be more important later, but after you have done a great job of steps a. through d., give some thought as to the role of third variables in your project.

   f. Keep things simple. Really understand your two variable relationships before you move on.

   g. Remember this is a process and your thinking will evolve as you work through these steps.

2. Research Design.

   a. What type of research design is best suited for project (surveys, aggregate data, observation, experiments, content analysis, etc.)? Think about the advantages/disadvantages of alternative approaches.

   b. Operationalization/measure major concepts: Consider reliability, validity and level of measurement. Develop appropriate measures of concepts and justify why they are good measures.

   c. Issues of sampling.

   d. Other concerns that come up in designing your project.

3. Data analysis.
a. Univariate statistics, particularly on the dependent variable and discussion of what is learned from them. Raw data, frequencies, central tendency, dispersion, etc.

b. Bivariate statistics which test your hypotheses.

c. Later on, multivariate statistics which test rival hypotheses.

4. Conclusions/implications

a. Is the hypothesis supported or not? Explain.

b. What has been learned about the hypotheses and what are its political/social/policy implications?

c. What has been learned about the methods you used?