IR 611: Multivariate Analysis for International Studies  
Professor: Benjamin Graham  
Course Meetings: Thursdays, 2:00-4:50pm  
Course Location: VKC 204  
Office: VKC 357B  
Version: January 8, 2020  

Office Hours: By appointment only. M 2:30-3:30; T&Th 10:50-11:20  
Sign up to meet with Professor Graham here:  
https://benjamingraham.youcanbook.me/  

Stats Consulting Office Hours: Times TBD.  
Sign up for statistical computing help here: https://statsconsulting.youcanbook.me/  

Learning Objectives:  
By the end of this course, students should be able to:  
1. Read quantitative research intelligently, identifying key assumptions, strengths, and weaknesses of quantitative hypothesis tests.  
2. Master several basic techniques for quantitative hypothesis testing, particularly ordinary least squares regression  
   a. Students will know the Gauss-Markov assumptions and understand how to proceed in the event that one or more of these assumptions is not met.  
3. Understand the state of the field well enough to engage in further self-teaching and work with co-authors to employ more sophisticated techniques than those developed in this course.  
4. Develop a working knowledge of R to facilitate goals two and three, including the ability to clean and merge datasets, run basic regressions, and create publication-grade regression tables and figures.  

The first part of the course focuses on the derivation, estimation, and application of linear regression. The second part of the course covers violations of the assumptions of the linear regression model and methods that are used when these assumptions are violated.  

The course will cover both the derivation and properties of estimators, as well as practical applications through data analysis using R, a statistical software package. The goal of the course is to provide students with an introduction to the tools that are necessary to conduct quantitative research in political science.  

Required Texts  
- I will also use a couple of chapters from the following book, but I will provide those chapters to you as a PDF. However, this is a good book to have on your shelf for
I will also post additional course readings in our shared google drive folder.

There are many other excellent textbooks and articles that address the topics from the course in greater detail. Sometimes, it also simply helps to have the same thing explained by a different author. The following is a list of sources that you can rely on to supplement the required books:


Course Support
This course is supported by the Security and Political Economy (SPEC) Lab. See the link to stats consulting office hours above. The SPEC Lab director, Alix Ziff, will also be running regular discussion sections for the course.

Students with Disabilities
Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. The phone number for DSP is 213-740-0776. If you need assistance with the process, please contact me at the phone number or e-mail address above.

Academic Integrity
I take academic integrity seriously. Any student violating USC’s Academic Dishonesty or Student Conduct policies will earn an ‘F’ in the course and will be reported to the appropriate administrators for investigation. You should review the Trojan Integrity Guide: http://www.usc.edu/student-affairs/SJACS/forms/tio.pdf and the guide for avoiding plagiarism: http://www.usc.edu/student-affairs/SJACS/forms/tig.pdf
Assignments

Quizzes: 5%

Problem sets: 40%
We will have a variety of homework assignments. Assignments will consist of mostly of problem sets related to regression, and working with data sets in R. Homework will be assigned in class and will be due at the beginning of class on the due date. Late homework will not be accepted. If you do not turn in an assignment on time, you will receive a zero on that homework.

Reproduce and Extend paper: 50%
Students are required to submit a final paper on the last day of class. In the paper, students should use the statistical skills learned in class to reproduce the results of a prominent paper in their field and then probe the robustness of that finding. Students will select a paper to reproduce in conversation with Professor Graham. Robustness tests may include the use of alternative measures for either the dependent variable or the independent variable of interest, adding additional control variables, modifying the sample, exploring interaction effects, or a variety of other modifications to the regression model. Papers should be double-spaced and 15-25 pages in length (including tables, graphs, and references). I will discuss more details about the paper as the semester progresses.

Student presentation: 5%
Students will present (and give feedback on) the results of their replication papers during the last session of the semester. Presentations will be less than 10 minutes in length (5-10 slides) and should include a summary of: theory and hypotheses, research design, comparison of original and replicated results, and discussion of potential issues. The final submitted version of the replication paper should incorporate the feedback received during this session.

Schedule of Topics: Readings my change as the semester progresses
Week 1 (January 16)-
1) Course Overview:
2) The Research Design Context
3) The Empirical Paper Checklist (Markowitz)
4) Re-introduction to R

Due: Before this class meeting, please download and install both R and R studio on your computer.

If you need it: Alix Ziff will offer an “R basics workshop” prior to our first course meeting in which she will help you load R and R studio onto your computer, figure out how to set your working directory, and get you acclimated to R in general. This
workshop is strongly recommended if you missed math bootcamp or need a ground-up review.

Handed Out: Homework 1: R basics & Data Investigation

**Week 2 (January 23) – Probability and the Structure of a Quantitative Paper**
1) Let’s talk about the Jensen paper!
2) Probability introduction
3) Set notation
4) Conditional Probability
5) Introduction to data management in R
   - Focus on variable manipulation with *dplyr* commands

Required Reading:
1) Nathan Jensen, *Democracy and FDI*
2) QSS Chapters 6.1 – 6.2
3) Jim Vreeland, “Goldilocks and the Three Regimes” Look at *slides 7-21*.

Handed Out: Homework 2: Data Management I – Manipulating Variables
Due: Homework 1: R basics & Data Investigation

**Week 3 (January 30) – Probability & Distributions**
- Random variable
- Uniform and normal probability distribution
- Expected value and variance of a random variable
- Law of large numbers
- Central limit theorem
- Data Management II
  - Focus on reshaping and merging datasets with *tidyr* commands

Required Reading: QSS Chapters 6.3

Handed Out: Homework 3: Data Management 2 – Reshaping and Merging Datasets
Due: Homework 2: Data Management 1

**Week 4 (February 6) – Uncertainty and Hypothesis Testing**
- Bias and mean squared error of point estimators
- Confidence intervals
- One sample tests & two sample tests
- Interpreting p-values
- Group exercise: Working with experimental data

Required Reading: QSS Chapters 6.3, 6.4 and 7.1
Due: Homework 3: Data Management 2

Next Week: I will run extra office hours to discuss your choice of paper to reproduce and extend.

**Week 5 (February 13) – Linear Regression I**
- T-tests vs. Z-tests
- The method of least squares
- linear statistical models
- Exercise: Can you draw the regression line?

Required Reading: GP Chapter 3

Recommended Exploration: Mess around on this website for awhile – it is a good way to build intuition of how ordinary least squares (OLS) regression works. http://trialandstderror.com/mechanics.html

Handed Out: Homework 4: Data Management III – Application to real data
Due Today: Email me the paper you want to reproduce and extend for your final project.

**Week 6 (February 20) Linear Regression II**
- assumptions of linear models
- Gauss-Markov Assumptions
- Goodness of fit
- confidence intervals for regression coefficients
- F test

Required Reading: QSS Chapter 7.3

Due Today: Homework 4: Data Management III

**Week 7 (February 27) Multiple Regression**
- estimation and interpretation
- confidence interval
- p-value
- How to make a regression table

Required Reading: GP Chapters 5 and 7

Quiz Today: Write down all the Gauss Markov Assumptions
Handed out: Homework 5: Making Regression Tables
Due today: Email me a paper you think we should read for class in Week 9.

**Week 8 (March 5) Multiple Regression II**
- interaction effects, polynomial regression
- log-linear regression
- scaling and units of measurement
- Data Visualization in ggplot2

Required Reading: GP Chapters 6 and 8
Recommended: GH Chapter 4
Due: Homework 5: Making regression tables

**Week 9 (March 12)** Reading and Reviewing of Empirical Papers
- Let’s talk about those papers!
- More data Visualization in ggplot2

Required Readings (TBD)
Due: You have prepped your data for your replication paper
Handed out: Homework 6: Data Visualization I

SPRING BREAK

**Week 10 (March 26)** Simulation
- analytical computation of marginal effects in regression models
- simulation of probability models
- summarizing linear regressions using simulation

Required Reading: GH Chapters 4 & 7 (skip 7.4)
Due: Homework 6: Data Visualization I

**Week 11 (April 2)** Multicollinearity and Heteroscedasticity
- detecting heteroscedasticity
- robust standard errors
Required Reading: GP Chapter 10 & 11

**Week 12 (April 9)** Time-Series-Cross-Sectional Analysis
- detecting autocorrelation
- OLS estimation in the presence of autocorrelation
Required Reading: GP Chapter 12

**Week 13 (April 16)** Logit, Probit, and Multiple Imputation
Required Reading: GP Chapter 15


URL: [https://osf.io/preprints/socarxiv/46xc7/](https://osf.io/preprints/socarxiv/46xc7/), doi:10.1017/pan.2017.43
Week 14 (April 16)  Returning to the Big Picture OR Matrix Algebra and Regression by Hand
Readings TBD

Week 15 (April 30):  Student presentations of final projects

Supplementary materials available on the Internet:
An excellent introduction to statistics and research design is Statistics at Square One  
-- http://bmj.com/collections/statsbk/index.shtml, see especially Chapter 5 --  
http://bmj.com/collections/statsbk/5.shtml

Good websites on statistics, econometrics, including free downloadable  
software for data entry, data analysis, research design, hypothesis testing,  
document preparation and presentation include:
http://davidmlane.com/hyperstat/index.html
http://members.aol.com/johnp71/javasta2.html#Freebies

Online readings on the scientific method:
http://www.lse.ac.uk/collections/lakatos/
http://galileoandeinstein.physics.virginia.edu/lectures/lecturelist.html
http://teacher.nsrl.rochester.edu/phy Labs/AppendixE/AppendixE.html
http://plato.stanford.edu/entries/popper/
http://www.brint.com/papers/science.htm
http://www.emory.edu/EDUCATION/mfp/Kuhnsnap.html
http://wwwcdf.pd.infn.it/~loreti/science.html