Political Data Analysis – POS 6737 Section 7129
Fall 2011
R: Periods 2-3; Room: MAT 102
R: Period 4; Political Science Data Lab

Department of Political Science
University of Florida

Instructor:
Prof. Badredine Arfi
Office: 221 Anderson Hall
Phone: (352) 273-2357
Email: barfi@ufl.edu
Office Hours:
Thursday 11:00 - 12:00
Friday 10:00 - 12:00

Teaching Assistant:
Brian Amos
Office: 001 Anderson Hall
Phone: (607) 227-8965
Email: 2011DataAnalysis@gmail.com

Course Description and Objectives

This course introduces students to the theory and techniques of statistical data analysis. At the end of the semester diligent students should find themselves well equipped with the tools necessary to develop their own statistical models for data analysis. The course has three main goals. First, the students are expected to learn how to design and carry out research that employs statistical techniques as a means of testing substantive theories of politics. Second, the students are expected to build a good foundation in statistics that would prepare them for learning more advanced statistical tools and analysis. Third, the students are expected to learn enough statistical skills to be able to understand as well as engage published works in political science research that utilize statistical analysis as means of testing theoretical arguments. In the weekly class meeting the course will more or less be conducted as a lecture-based workshop. The bulk of learning in the course, however, will take place outside of the classroom. Labwork will constitute an important component of the learning enterprise – learning how to use available statistical software – Stata 11 – is a must to succeed in this course.

Requirements and Evaluation

The requirement for this course is simple: work diligently and persistently. This includes attending classes, doing the readings carefully before the seminar meets, and working regularly on the problem sets and the research paper. Each student should expect to be spending many hours learning how to effectively use the statistical Stata 11 software commonly used to estimate the models discussed in class.

There will be a number of homework assignments that the students must complete and turn in. The homework assignments are due on the specified dates; no late submission is acceptable. In addition, the students are strongly encouraged to solve the odd-numbered exercises at the end of each chapter of the Agresti textbook. The answers for the Agresti problems are provided at the back of the book. This is a powerful way to put into practice the concepts learned in each chapter as well as provide you with much needed exercise to effectively understand and master the purported statistical skills.

A major component of the course evaluation will be a term research paper. Each student will produce a manuscript of high quality using an appropriate modelling strategy. Both my assistant and I will work closely with the students on their projects.

Distribution of Grades

- **25%:** 5 weekly graded homework assignments from Agresti Book: each will count for 5% of the final grade. Overall the homework assignments will count for 25% of the overall grade. No excuse will be accepted for not turning any assignment (except when justified with officially acceptable documentation). All assignments are due typed and double-spaced at the beginning of class on their respective due dates. No late submission accepted for any reason (except when justified with officially acceptable documentation).

- **25%:** 5 weekly graded homework assignments from Stata Book: each will count for 5% of the final grade. Overall the homework assignments will count for 25% of the overall grade. No excuse will be accepted for not turning any assignment (except when justified with officially acceptable documentation). All assignments are due typed and double-spaced at the beginning of class on their respective due dates. No late submission accepted for any reason (except when justified with officially acceptable documentation). These homework exercises are an excellent practice in learning how to use Stata 11

- **40%:** A research paper on a topic chosen by the student in consultation with the instructors. The goal is to produce a high quality manuscript, using a model (or models) discussed in the course. The research paper is due on the last day of classes and counts for 40% of the overall grade.

- **10%:** Practice of Stata 11 in Computer Lab for at least TWO hours every week counts for 10% of the overall grade. A record of practice will be held for each student. Every student is required to practice for at least 20 hours (i.e., 10 sessions) during the whole semester. A completed 2-hour practice is worth 1% of the overall grade (10 sessions 10%). No excuses will be acceptable for any reason. If you miss the practice for any one week you must make it up in another week, but you cannot have more than two sessions per any one week.

Required Reading Materials

- Agresti, Alan and Barbara Finlay. 2008. *Statistical Methods for the Social Sciences*, 4th Edition. Upper Saddle River, NJ: Prentice-Hall.

- Acock, Alan C. 2008 (3nd Edition). *A Gentle Introduction to Stata*. College Station, TX: Stata Press.

Computer Requirements

All models covered in this class can be estimated using Stata 11 software package. It is a must that you have a computer account and password so that you can use the computers in the Anderson datalab. Stata 11 is available on all of the Anderson Hall datalab computers.

Research Paper Strategy

Each student must:

- Identify a significant research question in his/her field of study
- Choose data from a data set which represents the variables involved
- Conduct analysis to address the research question using one of the techniques discussed in the course.
- The final product should be 12-15 pages long, include statistical analysis, and bibliography.

In order for the instructors to provide guidance in the preparation of this paper, you will be required to turn in various brief intermediate papers throughout the semester. Always submit 2 copies of each assignment – one electronic file (**whenever it is not impossible**) and one hard copy. All assignments should be handed in to my teaching assistant (hard and electronic copies emailed to the teaching assistant).

DUE DATES FOR PAPER ASSIGNMENTS		
Phase I	**September 22**	• Topic of the paper
Phase II	**October 6**	• Research question and hypothesis • Extended substantive bibliography
Phase III	**October 20**	• Description of the data to be used • How and when it is to be secured
Phase IV	**November 10**	• Description of the data analysis technique to be used • How it is to be applied in this case • Discussion of why it is the appropriate approach.
Phase V	**December 1**	• Data analysis results done to address the question
Phase VI	**December 7**	• Two copies of the completed paper – One electronic and one hard copy

WEEKLY READINGS, HOMEWORK ASSIGNMENTS AND OUTLINE OF THE COURSE			
Week	**Date**	**THEORY TOPIC**	**Agresti Chapter**
1	8 – 25	• Introduction to the Technology of Data Analysis through Statistics and visit to the computer lab.	1
2	9 – 1	• Sampling and Measurement	2
	HMWK 1	• 1.8, 1.14, 2.4, 2.24	
3	9 – 8	• Descriptive Statistics	3
4	9 – 15	• Probability Distributions	4
	HMWK 2	• 3.20, 3.48, 4.20, 4.38	
5	9 – 22	• Statistical Inference: Estimation	5
6	Thanksgiving		
7	9 – 29	• Statistical Inference: Significance Tests	6
	HMWK 3	• 5.24, 5.40, 6.24, 6.32	
8	10 – 6	• Comparison of Two Groups	7
9	10 – 13	• Analyzing Association between Categorical Variables	8
	HMWK 4	• 7.18, 7.36, 8.16, 8.22	
10	10 – 20	• Linear Regression and Correlation	9
11	10 – 27	• Introduction to Multivariate Relationships • Multiple Regression and Correlation	10 & 11
	HMWK 5	• 9.16, 9.24, 10.14, 11.22	
12	11 – 3	• Comparing Groups: Analysis of Variance Methods	12
13	11 – 10	• Combining Regression and ANOVA: Analysis of Covariance	13
14	11 – 17	• Logistic Regression: Modeling Categorical Responses	14
15	12 – 1	• Logistic Regression: Modeling Categorical Responses	15

WEEKLY READINGS, HOMEWORK ASSIGNMENTS FROM STATA BOOK			
Week	**Date**	**COMPUTER PRACTICE TOPIC**	**Acock Chapter**
1	8 – 25		
2	9 – 1		
3	9 – 8	• Getting Started	1
	SHMWK 1	• 1.1, 1.2, 1.3, 1.4	
4	9 – 15	• Entering Data	2
5	9 – 22	• Preparing Data for Analysis	3
	SHMWK 2	• 2.2, 2.4, 3.1, 3.6	
6	Thanksgiving		
7	9 – 29	• Working with Commands, do-files, and Results	4
8	10 – 6	• Descriptive Statistics and Graphs for a Single Variable	5
	SHMWK 3	• 4.1, 4.2, 4.3, 5.1, 5.6	
9	10 – 13	• Statistics and Graphs for Two Categorical Variables	6
10	10 – 20	• Tests for One or Two Means	7
	SHMWK 4	• 6.1, 6.4, 7.2, 7.5	
11	10 – 27	• Bivariate Correlation and Regression	8
12	11 – 3	• Analysis of Variance (ANOVA)	9
	SHMWK 5	• 8.4, 8.6, 9.3, 9.6	
13	11 – 10	• Multiple Regression	10
14	11 – 17	• Logistic Regression	11

Links to Some Data Archives

ICPSR – Inter-University Consortium for Political and Social Research
ANES – American National Election Studies
ARDA – America Religion Data Archive
U.S. Census – Census statistics for 2000
Center for the Group Study of Processes – Social Psychology
Africa Research Program – Studies on the economy of African States
CSAE – Center for Studies of African Economies
IRLA – Internet Resources for Latin America
CIAT – International Center of Tropical Agriculture
World Bank – World Bank datasets
IPUMS - Integrated Public Use Microdata Series (paid - international and domestic)
http://ssdc.ucsd.edu/index.html