EPPS 7316
Regression and Multivariate Analysis
Spring 2016, Tuesday 4:00-6:45, SLC 1.202

Prof. Patrick T. Brandt
E-mail: pbrandt@utdallas.edu
Web: http://www.utdallas.edu/~pbrandt
Office: Green 2.802
Hours: Tuesday and Thursday 12:00-1:00pm; Thursday 2:30-4:00pm; and, by appointment

TA: Lola Arowolo
E-mail: Olufunlola.Arowolo@utdallas.edu
Office: Green 2.816
Hours : Tuesday 1:00-3:45pm, Friday, 9:00am-12:00pm, and by appointment

Description

This is a graduate lecture course that covers multivariate data analysis and advanced regression applications for political economy, political science, criminology, and public policy students. The goal of the course is to ensure that graduate students in these fields are able to interpret, implement, and employ basic regression analysis of social science data.

The course is primarily devoted to advanced regression analysis. The emphasis is on developing a rigorous, calculus and matrix algebra-based understanding of ordinary least squares (OLS), violations of regression assumptions, hypothesis testing, and the general linear model. The emphasis is on using regression models to test social and economic hypotheses. Most of the course will be spent developing basic approaches to regression analysis when the assumptions of OLS fail. Some time will be spent on estimators other than least squares when assumption violations require these methods. The main focus is regression models, since these are the basis for most other methods for analyzing multivariate data. The course ends with an introduction to advanced topics in regression analysis, qualitative response models, and non-OLS approaches to estimation. Topics are supported by computer-supported data analyses using application-specific software

The emphasis of this course is on the acquisition and understanding of *analytical techniques*. It is very much an applied course, but for applications to be informed, there is a fundamental level of theoretical knowledge required. Many mistakes will be made by those who only know how to interpret findings and use "canned" computer packages. We emphasize that students know exactly what is going on when they do statistical analysis.

Student Learning Objectives

On completing this course, students will be able to:

- Apply and develop linear statistical models using political, economic, and social science data
- Recognize the basic assumptions, as well as strengths and weaknesses, of models using multivariate analyses of relevant social science data
- Acquire, prepare, and code data relevant to the student's research area and,
- Implement and interpret multivariate analysis in their own independent research project.

Required texts and course materials

Texts

Four books have been ordered. They are:

Required:

Gujarati, Damodar and Dawn Porter. 2009. *Basic Econometrics.* 5th Edition.

Kennedy, Peter. 2008. A Guide to Econometrics. 6th edition. Blackwell Publishing

Gill, Jeff. 2006. *Essentials Mathematics for Political and Social Research* Cambridge: CUP (you may have this from last term).

Fox, John R. and Sanford Weisberg. 2011. *An R Companion to Applied Regression*. Sage

Another new text you may wish to consult is

Monogan, Jamie. 2015. *Political Analysis Using R*. Springer.
`http://link.springer.com/book/10.1007/978-3-319-23446-5`

Students find the Kennedy book to be extremely helpful, so I encourage students to buy it. In general you should use all the books as reference materials. If I have not indicated a reading from a book it does not mean that a topic is not covered in it. Thus, use the texts as reference materials for each week and take the indicated readings as a guide. It is my expectation that you will find additional texts and resources to supplement the assigned course materials. Some sources will be indicated in course lectures and notes.

Students will be expected to use other advanced sources for data and statistical reference, including JSTOR, journal databases, Statlib, POLMETH, and other resources to which you will be directed.

As for software and applications, we will be using both Stata and R. I know you have already seen some things in Stata and I will illustrate concepts in both software packages as appropriate. The Fox text is considered one of the best for getting started in R in a regression context.

eLearning

eLearning is used in this class for distributing course materials (notes, datasets, code, etc) and to check your progress in the class. If you prefer to work on campus, you may do this at the library or in the many computer labs on campus. You can access the course eLearning page at `http://elearning.utdallas.edu`. Additional information about how to use eLearning is available at this site as well. You will need a UTD net-id to access this site. eLearning is also how I will communicate with you. You are responsible for announcements made through eLearning.

Lecture notes

I will make available copies of my lecture notes / slides via eLearning. I will typically provide these in advance of covering the material.

Computing and Course Materials

A major component of applied statistics is using computers and data to implement models and test hypotheses. Toward that end, students should be expected to utilize a variety of statistical packages for their work. While no one statistical program will be used for this course, familiarity with several is important because each has their own niche. I will primarily use:

Package	Cost
R	Free
Stata	Available at UTD or with a student discount

Any others are acceptable, and you should *not* express reservation about learning additional programs — it is an impediment to your research and teaching.

For home use (i.e., those who hate to purchase licenses), I highly recommend R, the GNU clone of S-plus. It is free, has great graphics and is documented well (mostly). You can get started in setting up and using R via Chapter 1 of the Fox text. I would suggest you do this immediately after the first class session and start to work through the early examples in that text.

I generally present examples in class using Stata or R, whichever is more appropriate to illustrate the material. I will not spend a great deal of time leading tutorials on statistical software in class. I am happy to have them arranged outside of class. I will also make programs and code available on an as needed basis in several statistical packages to demonstrate techniques. Note that somethings are easier to do in different packages and then convert the data to another format for analysis (I often construct variables and datasets in Stata and then move them to other software for analysis.) Your best resource for learning and implementing new methods are your peers, the TA, and the voluminous manuals that come with statistical software!

If you are looking for help with the software, the following websites will be of interest:

Stata Good help site are

- UCLA site: `http://www.ats.ucla.edu/stat/stata`
- Princeton site: `data.princeton.edu/stata`
- Official on-line Stata help: `http://www.stata.com/capabilities/search-help-files`

R Good starting places (other than Googling "R" or "CRAN") are

- UCLA site: `http://www.ats.ucla.edu/stat/r`
- Rseek : `http://rseek.org`

Grading, Assignments, and Course Policies

Grading

There will be approximately eight to ten *graded* assignments, an exam, and an original research paper. The allocation of the grade is:

- Assignments: 25%
- Exam I: 25%
- Exam II: 25%
- Paper: 25%

The first exam will be in-class, approximately during week 8. The second exam will be a take-home exam at the end of the course.

Late assignments and papers will be penalized. Late assignments and papers will be penalized 10 points per day. Papers or assignments submitted via e-mail will be considered received by the date and time stamp on the e-mail received in my e-mail inbox. The paper will be due no later than April 30, 2016.

My very strong preference is to not allow incompletes in this course, since you need most of the material completed for later course work, the Methods Qualifying Exam, or your thesis / dissertation proposal. If

you believe you will need to take an incomplete you 1) must follow university policy, and 2) should contact me as soon as practicable (i.e., not the week of the final exam or the day before the final paper is due).

Grades are based on the standard grading scale: A = 100-90, B = 89-80, C = 79-70, etc. with plusses and minuses at my discretion.

If you have any questions about your grade on an assignment, please wait until 24 hours after receiving your assignment before discussing the grade with me. There are no exceptions to this policy.

If you wish to have an assignment re-graded, it must be returned to the me or the TA within two days of the day it was returned (if I am not available that day, ask the staff of the School of Economic, Political and Policy Sciences to leave it in my box). Assignments to be re-graded must include a memo stating the reason why you believe they assignment should be re-graded. Finally, re-graded assignments can be graded higher, lower, or the same as the initial grade.

Assignments

Assignments and exams will cover applied and theoretical problems. Assignments should be typewritten as much as possible. I realize that may require setting mathematical text or other typographical symbols. This can be done in standard word processing software. Feel free to only use Roman letters. If you are so interested, I can arrange a short tutorial on how to use LaTeX for this task.

Any statistical output or data analysis you do should be fully interpreted and presented as though it were being sent for publication to a journal. *This means that regression output from your statistical package of choice that is copied into a word processor document is unacceptable. You should take the time to typeset the results into a meaningful, well-labeled table or present a well documented and coherent graphical summary of any results.* If you have any questions about what to include in your data output and assignments, consult empirical work in standard journals (APSR, ISQ, AJPS, JOP, ASR, PAR, Criminology, etc.) or ask. Assignments that do not meet this requirement will be returned and not graded until revised. I will provide further guidance on this as the course progresses.

You may work together on assignments, but each person must turn in their own work. Working together has two benefits. First, it can help you see if you really understand the material (if you can explain it to someone else and convince them that you are right, you probably are). Second, it gets you in the habit of working with others to solve problems (and remember that most research is coauthored!) You do need to be careful about two issues in the course of working together. The first is plaigarism. The second is letting the person who "gets it" do all the work. Just because someone looks like they have solved a problem does not mean it is the correct answer. One of the things I have seen before is that working together can help on the "easy" problems, but on some of the harder problems, a "group" will often come up with the wrong answer. Beware of "groupthink"!

Finally, if you have questions about the assignments, I encourage you to come and ask me or the TA about them as soon as you realize you have run into a problem! Pounding your head on a desk for 6 days and then coming to get help on day seven (when the assignment invariably is due) is poor form and is not going to help you learn the material. One of the best methods I have found for asking and answering questions for this course is e-mail. The benefit of e-mail is that it forces you to compose your question(s) very specifically and to think through what you are asking logically. In addition, I can generally offer a faster response via e-mail than if you wait for office hours.

Papers

The paper will require students to analyze and interpret regression or multivariate model. The paper you write should be based on your own research and interests—there is no requirement to use a particular dataset or technique. However, the paper should be original (it should not include an analysis used in previous

courses, conference papers, or other writing unless we have specifically discussed it). The following are suggested models for a paper:

- A conference paper (examples can be seen on my webpage).
- A dissertation or thesis chapter (examples available upon request).
- A replication article (see the replication standard for Political Analysis) or `http://gking.harvard.edu/files/abs/replication-abs.shtml`

Before beginning your paper, you are **REQUIRED** to come and talk with me about your paper. At or prior to this meeting, please provide a ONE PAGE research design and lays out the following for your paper:

- Main research question you are addressing.
- Hypothesis (or Hypotheses) you wish to evaluate in your paper.
- Data and variables to be used in the analysis.
- Tests and methods that will be used to evaluate the hypotheses.
- Tentative listing of the techniques and models you may use.

The data you use for the analysis is something that you should already have or have easy access to. Data collection is not a topic we are covering in this course and time spent building large complex datasets will detract from your ability to complete the paper adequately.

The meeting to discuss the paper should be scheduled before the 5th or 6th week of class (roughly the week of February 15th). Papers will be due on April 30, 2015.

Attendance

It should go without saying that in a class of this size your attendance is easily noted and therefore required. If you are unable to make a class or will be late, advise the instructor as far in advance as possible.

Course conduct

The following rules apply in class:

1. Turn off your cell phone. It is VERY distracting to others. "Off" means that it does not ring OR vibrate. I will ask you to leave if your phone rings. (Exceptions to this policy can be made, come to talk to me.)
2. Do not fall asleep. It is rude and distracting. Bring coffee if you need it (I do.)
3. Be polite and courteous to your fellow students.
4. Raise your hand when you want to be recognized to answer or ask a question. If you do not raise your hand I will not recognize you or your answer.
5. You are reponsible for things: therefore if you miss something I announced or fail to complete an assignment, my response will be "How is this my problem?"
6. Note that this syllabus is not a contract. It is subject to change at my discretion. While we may be studying the political science, this class is not a democracy.

7. Class starts at 4:00pm. Not 4:15pm. Be here on time.

8. Respect my time and I will respect yours. We are both busy.

9. *You are expected to be proactive in anticipating and planning for any absences or problems you will have in completing course work.* Make arrangements for possibly missed work prior to the due date is preferable and more likely to be successful than doing it after the due date.

Course Outline

Readings listed for each week are suggestions. My lectures and discussion will parallel these readings. The readings are listed in order of importance (from most to least important). Feel free to ignore or consult them as needed. I will regularly post or hand out lecture notes on the material we are covering.

In general, I have my own "order" to covering the material. This may differ from the texts, or what others might do. My experience is that this layout of the material works well.

Weeks 1-3 (1/12–1/26): Basic concepts: probability, what is linear regression, why do we want it, and other tools like matrix algebra

- Gill Ch 7-8
- Gujarati Ch 1-2, Appendices A and B
- Fox, Preface and Chs. 1–2

Week 4 (2/2): R and Stata

- Fox, Preface and Chs. 1–2

Week 5 (2/9): Regression estimation, three different ways: the likelihood principle, conditional expectations, and least squares

- Gill Ch 3-6
- Gujarati Ch 1-4, and 7
- Fox, Sections 4.1, 4.2, 4.8

Week 6 (2/16): "Tests and statistical insignificance": Inference in the regression model

- Gujarati Ch 5 and 8
- Kennedy Ch 4
- Fox, Sections 4.3–4.5
- Remember you should be talking to me by this week about your papers!

Week 7 (2/23): "Gaussian and the Gauss-Markov theorem may both be named for the same guy, but they are not the same": Properties of the classical linear regression model and the normality assumption

- Kennedy Ch 2-3
- Gujarati Ch 4

Week 8 (3/1): Specification decisions and extensions to the CLR

- Gujarati, Ch 6, 9, 13.
- Fox, Chapter 3
- Kennedy Chapters 5-6
- Gujarati Ch 13

Week 9 (3/8): Exam I

- Exam will be given in class this week.

Week 10 (3/15): Spring Break, no class
Week 11 (3/22): "Multiple measures on the same thing": Multicollinearity

- Gujarati Ch 10
- Kennedy Ch 12, 15

Week 12 (3/29): "It's a mad, mad, heterogenous, and correlated world": Generalized least squares, heteroscedasticity and serial correlation

- Gujarati Ch 11 and 12
- Kennedy Ch 8
- Fox, Chapter 6

Week 13 (4/5): "If you think you have one equation you are probably wrong.": Systems of equations, identification, non-fixed regressors

- Kennedy, Ch 9-11
- Gujarati Ch 13, skim Ch 18-20

Week 14 (4/12): "i and t, together forever": Panel data

- Gujarati Ch 16
- Kennedy Ch 18

Week 15 (4/19): "Sometimes our measurement is crude.": Limited and Qualitative Dependent Variables

- Gujarati Ch 15
- Kennedy Ch 16-17
- Fox, Chapter 5

Week 16 (4/26): "The best prediction of tomorrow is today plus a random shock.": Introduction to time series analysis

- Gujarati Ch 17, 21, 22
- Kennedy Ch 19-20

Exam II will be distributed the day the papers are due, 4/29/2016. It will be due via e-mail by 5pm on 5/6/2016. Details will follow later in the course.

University Policies

University course-related policies about course conduct, student discipline, academic integrity, e-mail, course withdrawls, grievance procedures, incomplete grades, disability services, and holiday policies are at `http://go.utdallas.edu/syllabus-policies`.

These descriptions and timelines are subject to change at the discretion of the Professor.