

**CRJ 604: Advanced Statistical Analysis  
Fall 2016**

**Instructor: Dr. Gary Sweeten**

**Office: UCENT 613**

**Telephone: 602.496.2342**

**Email: gary.sweeten@asu.edu**

**Office hours: 3:00-4:00 Tuesday & Wednesday, or by appointment**

**Class time: W 4:50-7:35**

**Class location: UCENT 256**

**Course Description**

This course will help you to become accomplished in the use of basic linear regression, the statistical workhorse for analyzing data. Researchers rely on this tool to estimate the relationship between variables so that they can describe correlations, reveal causal relationships, or make predictions. Ideally, estimates from this tool are unbiased (correct, on average) and efficient (close to the correct answer). Results from linear regressions possess these characteristics as long as certain assumptions can be made. When these assumptions cannot be made, certain fixes or different estimators can be applied. In this course we will investigate each of these assumptions: how to test them, and what to do in cases where they are not met.

**Course Objectives**

1. Students will understand the theoretical issues involved in the basic linear regression model in its simplest form (bivariate regression) and multivariate form (multiple regression).
2. Students will also acquire fluency with the computer application (using Stata) of bivariate and multivariate regressions and probit/logit models, including testing assumptions and applying fixes.

**Required Text**

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach, 4e*. Mason, OH: South-Western.

**Other Useful Texts**

Berry, William D. 1993. *Understanding Regression Assumptions. Quantitative Applications in the Social Sciences, Vol. 92*. Newbury Park, CA: Sage.

Cameron, A Colin and Pravin K. Trivendi. 2010. *Microeconometrics Using Stata: Revised Edition*. College Station, TX: Stata Press.

Fox, John. 1991. *Regression Diagnostics. Quantitative Applications in the Social Sciences, Vol. 79*. Newbury Park, CA: Sage.

Hardy, Melissa A. 1993. *Regression With Dummy Variables. Quantitative Applications in the Social Sciences, Vol. 93*. Newbury Park, CA: Sage.

Kennedy, Peter. 2008. *A Guide to Econometrics (6<sup>th</sup> Ed.)* Cambridge, MA: Wiley & Blackwell.

- This book is very useful for developing intuition. Earlier (and cheaper) editions will prove equally useful to this end.

Pampel, Fred C. 2000. *Logistic Regression: A Primer. Quantitative Applications in the Social Sciences, Vol. 132*. Newbury Park, CA: Sage.

### **Course Requirements**

Your course grade will be based on your performance on homework assignments, two exams, and one project.

You may cooperate on homework assignments; however, each student must turn in a separate assignment, and each student is responsible for the content of that assignment. To receive full credit, you must show your work and include your Stata output. All homework assignments are to be handed in (preferably emailed) by 6:00 pm on the due date. Late homework assignments will be docked 5% of the total points each day they are late.

The midterm will be a take-home exam. You may use your notes and the books for this exam, but you may not cooperate with other students on it. The final exam (also take-home) will focus on the material since the midterm exam. However, understanding of this material will require you to integrate knowledge from the earlier part of the course.

Each student will complete a project which demonstrates knowledge of the methods learned in class. This paper will take the form of an empirical section of a research paper (data, methods & discussion). The intention is that you will actually work with the data that you intend to use in your thesis, dissertation, or other research project. Of course, this might prove impossible. You may need to find another related dataset to analyze. If the dataset you are using for your thesis/dissertation is very large and/or complicated, it is appropriate to use a small subset of the data, or a related, more accessible dataset. The point of the project is not to finish your research project, but rather to get you started, and give you some practice working with data while answering a research question. Your project may take the form of a critical replication of an existing study using either the same dataset or a similar dataset. The due dates for each segment are listed below.

Part 1: Proposal (10% of project grade, due September 14)

2 page description of the data including a short description of the research question, identification of the data source, current status of data (e.g., in your possession, ready to analyze, or nowhere to be found), an enumeration of the sample (number of observations), a description of the strengths and weaknesses of the sample relative to the problem at hand, and a list of the relevant variables including their scale. The dependent variable can be either continuous or binary and you should have a minimum of 30 observations. The problem should be stated in terms of causality where you are interested in the impact of ONE independent factor on ONE dependent variable. (Bear in mind, however, that you most likely will not be able to claim that your estimated effect is in fact causal.)

Part 2: Final Paper (70% of project grade, due November 23)

Your final project should be a truncated research paper containing a very brief front end, and a full data, results and discussion section. There is no minimum or maximum length, but it should be shorter than the average journal article length in *Criminology*. Use *Criminology*'s [author guidelines](#) to format your paper.

The central analysis should be a multivariate regression with hypothesis test(s) on the variable(s) of interest. This paper should contain an abbreviated (no more than two pages) "front end" describing your research question, previous research on the topic, your theoretical framework and hypotheses. After this point, your paper should proceed as a typical journal article would, with data, results, and discussion sections, but not a detailed conclusion section. Your discussion section should contain a section testing the adequacy of your models using methods covered in class. Your discussion section should include ways to improve the analysis with alternative methods. This is worth 70% of your project grade. Your grade on this paper will be based on the extent to which it achieves:

- 1) a succinct, descriptive title and abstract (5%);
- 2) a clear and concise "front-end" with a statement of the research question, a short summary of previous research, a brief outline of the theoretical perspective, and a hypothesis/hypotheses (15%);
- 3) clear and detailed description of the data source, sample size, and variables drawn from the dataset (20%);
- 4) an accurate and clear discussion of the methods employed in the project (15%);
- 5) clearly presented findings including appropriate figures and tables as necessary (25%);
- 6) a reasoned discussion section that places the results in the context of the previous literature and the research question, and suggests ways to improve the analysis in future work (10%);
- 7) correct formatting, spelling, and grammar (10%).

Part 3: Presentation (20% of project grade, November 30)

You will present your project to the rest of the class in the format of an ASC panel presentation. You will have 10 minutes to present your project, followed by up to 5 minutes to answer questions. Your presentation will be graded on organization (15%), effective framing of the research problem (25%), clear communication of results using tables and/or figures (25%), demonstration of knowledge in answering questions (15%), and correct formatting, spelling, grammar, and timing (20%).

Components of final grade

|          |     |
|----------|-----|
| Exams    | 50% |
| Homework | 25% |
| Project  | 25% |

| <u>Percentage earned</u> | <u>Grade</u> |
|--------------------------|--------------|
| 98-100                   | A+           |
| 92-97                    | A            |
| 90-91                    | A-           |
| 88-89                    | B+           |

|       |    |
|-------|----|
| 82-87 | B  |
| 80-81 | B- |
| 78-79 | C+ |
| 70-77 | C  |
| 60-69 | D  |
| <60   | E  |

### **Students with additional needs**

I will make the necessary accommodations for any student with special needs documented by the Disability Resource Center. If this is the case, please inform me as soon as possible so that I can adequately prepare.

### **Academic Integrity**

All students are expected to adhere to the highest standards of academic integrity. Students should familiarize themselves with the ASU Student Academic Integrity Policy available online: <http://provost.asu.edu/academic-integrity>. This policy details behaviors considered to be violations of academic integrity and provides guidelines on the imposition of sanctions, such as grade reduction, suspension and expulsion.

**Tentative Course Outline** (subject to change, readings for that day listed below topic)

#### **8/24 Introduction and review**

Wooldridge, Appendices A-C

#### **8/31 Bivariate regression**

Wooldridge, Chapters 1 & 2

#### **9/7 Carrying out an empirical project / significance testing**

Wooldridge, Chapter 19, Appendix C.6

Bushway, Sweeten & Wilson 2006

Greenland et al. 2016

#### **9/14, 9/21 & 9/28 Multiple regression estimation and inference**

##### **Project proposal is due Sep 16th**

Wooldridge, Chapters 3 & 4

#### **10/5 Take-home midterm due; Further issues in regression analysis**

Wooldridge, Chapters 5 & 6

#### **10/12 Regression with qualitative information**

Wooldridge, Chapter 7

#### **10/19 Heteroscedasticity**

Wooldridge, Chapter 8

#### **10/26 Specification issues**

Wooldridge, Chapter 9

**11/2 Limited dependent variables**

Wooldridge, Chapter 17

**11/9 To be determined**

**11/16 American Society of Criminology Annual Meeting (no class)**

**11/23 Day before thanksgiving (no class); final paper due**

**11/30 Presentations**

**12/7 take-home final due**