Course Description

This course provides an introduction to the theory, methods, and practice of multivariate analysis. The course will not dwell on statistical theory. But, we will concentrate on: The importance of evaluating empirical relationships between variables as a component of the theory-testing process; the utility of regression analysis for doing so; the nature of the basic regression model and several extensions to the model.

Course Goals
The goal is to provide students with the skills that are necessary to: (1) read, understand, and evaluate the professional literature that uses multivariate analysis; (2) design and carry out studies that employ techniques for testing substantive theories; and (3) prepare to learn about more advanced statistical procedures.

Grading

- Class and lab attendance are required. Students missing more than two meetings will not receive a passing grade.
- Data preparation assignment (April 8) – 10% of final grade
  - For this assignment you need to present a dataset you downloaded from the internet and prepared for analysis. You need to submit the prepared datafile with 10–20 variables of interest, and a document explaining what the data are (who collected, how, when and any additional information that may be relevant), a list of all variables and how they were coded, and summary statistics of each variable of interest. Submit the paper along with your do-file.
- Four (4) Data Analysis Assignments (date varies) – 50% of final grade
  - Submissions should include the word file, dataset and a self-executing dofile. All variables should be labeled and coded appropriately. The data should be self-explanatory when opened. All files should be uploaded into Moodle.
  - In each of the data analysis assignments, you are required to discuss briefly your theory and hypotheses, data used, present descriptive statistics for the main variables of interest, discuss the model chosen and why it is appropriate for your data and argument, analyze the data using the model mentioned, examine your results and discuss your findings. Your specification of the model needs to confirm with existing work in your field of research. You should therefore refer to existing work in your field of interest. These papers should not exceed 4 pages (including graphs and figures).
  - You need to submit at least four (4) data analysis assignments. All assignments (combined) amount to 50% of final grade, each is weighted equally.
  - The topics of the assignments are listed below. For each assignment, we will post a detailed requirement list. We may not get to each one of these models but you may do them after consulting with us.
    1. Analysis of means and variances
    2. OLS Regression
    3. Logistic Regression
    4. Categorical (Multinomial or Ordinal) Regression
    5. Regression using a count dependent variable
    6. Moderation, Random Effects or Mediation (you may submit only 1 of these)
• Research note - 40% of final grade  
  ◦ Submit a 10 page research note of your research, which includes theory, hypotheses, data, analysis, and summary of findings. You may use any (or none) of the models you estimated in your data analysis papers. Make sure to incorporate the comments you received on the relevant data assignment papers. If you use a different model than those used in the data analysis assignments, explain why. Think of this as a research paper. Pose an argument, select data, and try to provide the best evidence you find for your argument (null findings are fine too).  
  ◦ Submissions should include the word file, data file (dataset) and self executing dofile. All variables should be labeled and coded appropriately. The data should be self-explanatory when opened. All files should be uploaded into Moodle.

All papers should be submitted according to conventional academic guidelines.

Plagiarism will be handled with accordance to the IDC bulletin and policies.

---

Learning Outcomes

• Data
• Comparing Means (t)
• Analysis of Variance (ANOVA)
• Multivariate Regression Models  
  ◦ Ordinary Least Squares (OLS)  
  ◦ Binary Dependent Variables (Logit)  
  ◦ Categorical Dependent Variables (Multinomial and Ordinal Regression)  
  ◦ Count Dependent Variable
• Regression Topics  
  ◦ Moderation  
  ◦ Random Effects  
  ◦ Mediation
• Factor Analysis (Exploratory & Confirmatory)

---

Lecturer Office Hours

Office: Lauder O127 (Lauder courtyard)

Office hours: Thursday, 16:00-17:00, and by appointment
Tutor Office Hours

- Using our statistical software
- Means
- ANOVA
- OLS
- Binary Logistic Regression
- Multinomial and Ordinal Regression
- Count Models
- Moderation
- Random Effects
- Mediation
- Factor Analysis

Additional Notes

The course assumes basic knowledge of statistics, statistical programming and research methods for social science at the graduate level. Students without this background need to take the course offered by the program before taking this course.

The statistical package we will use is Stata. You may use other software, but you are responsible for knowing how to use it for the requirements of this class. I recommend using Stata or R. Stata is available through the IDC. R is a free code.

Please note: Any use of electronic devices in class is prohibited. Students must silence and stow all electronic devices (e.g., laptops, iPads, cell phones, etc.) before class begins. Why don’t I allow such devices for taking notes? This post from Buzzfeed says it all: "11 Things You’re Actually Doing on Your Laptop During a Lecture."

Reading List

The reading list will be posted on Moodle.

For each class you are required to read one journal article that will be discussed in class. We will discuss these in class.

In addition, we will use the following book to review many of the models we cover in the course:


And, we will refer to the following book to make sure we understand what we can say and what we should not say when we use data:

The books are available from the IDC library.

We encourage you to supplement your work using web-based sources about statistics and about using Stata. We know you have the skills to navigate information from the worldwide web.