**DRAFT SYLLABUS**

[Last Modified: 4/25/2016]

**PPOL-G 745-01: Advanced Quantitative Methods**

**Spring 2016**

**Instructor Information**

Amit Patel, PhD

Office: McCormack Hall 3-425

Email: Amit.Patel@umb.edu

Phone (Cell): 718.866.5757 (Preferred)

Phone (Work): 617-287-5847

Office Hours: Wednesdays, 11:30 to 12:00 pm (after class), 2:00 to 5:00 pm, OR by appointment

**Course Information**

**Title:** Advanced Quantitative Methods

**Credit Hours:** 3

**Online Course?** NO

**Time:** 9:00 am – 11:30 AM, Wednesdays

**Classroom:** M03-0440 [PPOL/PAF Conference Room]

**Course Website:** <https://umb.umassonline.net> (BlackBoard)

**Pre-requisites:** PPOL-G 604/GERON-GR 603: Statistics I (or equivalent with permission of instructor), PPOL-G 605/GERON 604: Statistics II (or equivalent with permission of instructor)

**Course Overview**

This course introduces you to advanced quantitative research methods widely used in social sciences. The course is designed to give you a set of skills and knowledge about methods that you could apply in your dissertation research, at your workplace, or to prepare for more advanced courses in quantitative analysis. The course largely focuses on statistical methods but also introduces other quantitative approaches such as simulation methods and spatial analysis techniques. We will begin with the Ordinary Least Square (OLS) methods and then move on to the Maximum Likelihood (ML) methods and Generalized Linear Models (GLMs) that use ML estimators. The course then discusses the assumptions underlying the method of OLS (i.e., Gauss-Markov Theorem) and studies several detection methods and remedies for the violation(s) of the GM theorem. The course will introduce each of these methods with cross-sectional data to get the conceptual clarity first and then apply them to panel data for more robust analysis. The course stresses applications using statistical package, STATA, and introduces techniques to handle issues that are commonly encountered in social science research such as approaches to deal with missing data, reducing number of variables from survey data into single variable using factor analysis etc. The course will also introduce techniques to select best regression models within a method using techniques such as step-wise regressions and between methods (e.g. Fixed Effects vs. Random Effects or Poisson vs. Negative Binomial) using model selection criteria such as AIC and BIC. Towards the end of this course, we will have a brief introduction to several other quantitative research approaches such as Discrete Event Simulations, Queuing Models, and Agent-based Modeling. If there is interest, we could cover spatial statistical techniques such as point pattern analysis, location-allocation modeling, and spatial regressions.

The following topics could be covered in the course (final selection will be made on the first day of class based on participants’ preferences. The list is suggestive to facilitate final selection):

**Regressions with the Cross-sectional data**

* OLS Estimators
* Goodness-of-Fit Tests
* Dummy Variables and Interaction terms
* Multi-collinearity
* Prediction
* Maximum Likelihood Estimators (MLEs)
* Logit and Probit Regressions
* Other log-linear Regressions

**OLS Assumptions and the Violations**

* Gauss-Markov (GM) Conditions
* Tests for the GM Violations
* Weighted Least Squares (WLS) Estimation
* Generalized Least Squares (GLS) Estimation

**Regressions with Count and Censored Data**

* Poisson Regression
* Negative Binomial Regression
* Zero-inflated Poisson Regression
* Zero-inflated Negative Binomial Regression
* Tobit Regression
* Heckman’s Sample Selection Model

**Regressions with Time-Series, Pooled Cross-sectional, and Panel Data**

* Lagged Variables
* Difference-in-Differences Technique
* Fixed-Effects and Random-Effects Estimations

**Regressions with Endogeneity and Causal Mechanisms**

* Path Structure Analysis
* Instrumental Variables (IV)/2 Stage Least Square (2SLS) Estimations
* Simultaneous Equation Modeling

**Simulation Methods (Brief Survey)**

* Queuing Models
* Discrete Event Simulations
* Agent-based Modeling
* Micro-simulations

 **Spatial Statistical Methods**

* Point Pattern Analysis (Hot-spots, K-function)
* Location-allocation Modeling (P-median etc.)
* Spatial Regressions (Spatial Autocorrelation detection using Moran’s I and correction methods)

**Dealing with pragmatic issues in quantitative analysis (spread out during the semester):**

* Missing data
* Model selection methods (within method and between methods)

**Learning Outcomes**

**1. Knowledge and understanding:**

* Students should be able to read and understand the journal articles that involve advanced time-series/panel-data statistical analyses
* Students should be able to critically assess policy arguments, comments, reports and other materials that use advanced statistical analysis
* Where relevant, students should be able to make policy remarks and draw policy implications/conclusions based on the findings of various policy studies that apply statistical techniques.

**2. Develop quantitative and qualitative skills**

* Students should have skills to formulate social science research questions that could be answered with statistical methods covered in this course
* Students should be able to conduct time-series/panel-data statistical analyses to analyze various social and policy issues.
* Students should be knowledgeable about fundamental concepts and underlying logic of advanced statistical analysis
* Students be knowledgeable about strengths and limitations of inferences you could make from statistical analysis
* Students should have skills to select appropriate statistical methods for the data and policy questions at hand
* Students should be able to interpret statistical results through the STATA outputs of various statistical analyses covered under this course.
* Students should acquire basic programming skills to conduct time-series/panel statistical analysis in STATA.
* Students should learn how to find, collect, organize and clean datasets necessary to analyze the topic of their interests.

**3. Professional development and leadership**

* Students should learn working effectively in a team setting through homework assignment and course project.
* Students should learn how to professionally present statistical results to policy audience through a course project presentation.
* Students should learn how to seek assistance when needed and communicate effectively with classmates and instructor.
* Students should learn how to deliver assignments and outputs in a timely fashion.
* Students should learn formulating research question, collecting secondary data, conducting statistical analysis and present them verbally and in written format independently (skills required to carry out a doctoral dissertation successfully).

**Course Materials**

**Required Text:**

1. *Introductory Econometrics, 6th Edition* by Wooldridge, Jeffrey (Cengage Learning)

ISBN-10: 130527010X | ISBN-13: 9781305270107 + MindTap® Economics, 1 term (6 months) Printed Access Card ISBN: 9781337127141

**Recommended Text:**

1. Statistics with Stata - Updated for Version 9 by Hamilton (Duxbury Resource Center)

2. A Guide to Econometrics by P. Kennedy (MIT Press)

3. Applied Logistic Regression by D. Hosmer and S. Lemeshow (Wiley InterScience Publication)

4. Useful online statistical texts (to supplement the above texts)

<http://davidmlane.com/hyperstat>

<http://www.statsoft.com/textbook/stathome.html>

[http://www.itl.nist.gov/div898/handbook/](http://www.itl.nist.gov/div898/handbook/%20) (more technical than the above websites)

**Required Software:** STATA 14 (IC or SE)

I will recommend purchasing either a six month ($75) or one year license ($125) of **STATA/IC 14** through the GradPlan. The STATA/IC 14 could be bought at special price from this webpage: <https://www.stata.com/order/new/edu/gradplans/>

**Please DO NOT buy Small Stata; it has limitations on size of the dataset that it can handle.**

Alternatively use the computers on campus.

Buy Stata 14 SE or perpetual license for IC only if you anticipate multiple years of usage or plan to take statistics courses in future that use Stata.

We will be spending time using Stata in-class. Unfortunately, our class does not meet in a computer lab. **Please bring your own laptop computer for in-class work.**

## Class Preparation

PPOL-G 745 is a rapidly moving course. Preparing for each class will guarantee learning. Class time will be most productive if all of us have completed required readings and worked through the relevant examples before class. There may also be online material/tutorials posted each week. It is important that you review those materials before the class.

The required readings are consciously kept short to allow you more time to think about public policy applications of the methods you learn. You will learn more by puzzling over a problem than by reading chapters twice or memorizing formulas. Keep asking this question to yourself for each new statistical method you learn in this class: how can I use this statistical concept to analyze public policy issue or a social science research question of my interest? You will have a veto over ongoing activities in the class: I encourage you to stop me anytime during the class and simply ask, "How is this Relevant for Social Science Research?" We will stop right there and spend time on relevance of the concept for social science applications under discussion at that moment.

Some of the in-class activities require preparation before the class. We will not be able to move forward if you are not prepared and may waste everyone's valuable class-time. **As a guiding principle, for EACH hour you spend in the classroom, please set aside at least TWO additional hours in your weekly schedule.**

## Teaching Style

The class-time will be spent on three distinct type of activities: i) acquiring core of the statistical concepts through hands-on exercises; ii) understanding statistical concepts through lecturing; and iii) mastering software skills to implement statistical concepts with the actual data in STATA. Majority of the class time will keep you active; lecturing will be minimal. Your active participation is the key to achieve our learning outcomes.

## Grading Policy

Weekly Research Questions 10%

Online Participation 5%

Peer Learning Exercise 10%

Homework - 3 Problem-Sets (10% Each) 30%

Term Project (Written work: 5% + Class presentation 40%) 45%

|  |  |  |
| --- | --- | --- |
| **Letter Grade** | **Percentage** | **Quality Points** |
| A | 93-100% | 4.00 |
| A- | 90-92% | 3.75 |
| B+ | 87-89% | 3.25 |
| B | 83-86% | 3.00 |
| B- | 80-82% | 2.75 |
| C+ | 70-79% | 2.25 |
| C | 60-69% | 2.00 |
| F | 0-59% | 0.0 |
| **INC** | A grade of Incomplete (INC) is not automatically awarded when a student fails to complete a course. Incompletes are given at the discretion of the instructor. They are awarded when satisfactory work has been accomplished in the majority of the course work, but the student is unable to complete course requirements as a result of circumstances beyond his/her control. The student must negotiate with and receive the approval of the course instructor in order to receive a grade of incomplete. Incomplete grades are reserved only for major family or medical emergency. Documentation will be required. | N/A |
| IF | Received for failure to comply with contracted completion terms. | N/A |
| W | Received if withdrawal occurs before the withdrawal deadline. | N/A |
| AU | Audit (only permitted on space-available basis)  | N/A |
| NA | Not Attending (student appeared on roster, but never attended class. Student is still responsible for tuition and fee charges unless withdrawal form is submitted before deadline. NA has no effect on cumulative GPA.) | N/A |

## Weekly Research Questions

You will submit a research question that could be answered with the method covered in that week (Instructions for each week will be posted under Discussion Board section of the course website. This exercise will begin in **week 2**). Ideally, you should frame a question that is within your term-project topic area. The submission is very short: it includes a research question and description of what kind of data may be required to answer that question. It is NOT expected to actually attain the dataset and answer the question. The purpose of this exercise is two-fold: i) framing a question will allow you to understand the strength and limitation of the statistical method under study, and ii) you will advance your research for the term-project on weekly basis.

## Online Participation

There will be several opportunities to engage in a discussion with your peers. We will open new threads as questions come up during class discussion. There will be a perpetual opportunity to participate in online discussion: your colleagues’ weekly research questions. Please read and comment on at least one of your colleague’s contribution. Please share anything that is relevant for this course to your instructor or guest speakers (such as new data source to the librarian).

## Peer Learning Exercise

Goal of this exercise is to contribute to mutual learning experience. We will take turns and share a substantial Stata technique (e.g. ONE way of handling missing values), or share a major critique (e.g. the cult against use of statistical significance as a gold standard), or provide a walk-through of a resource (e.g. how to obtain dataset from ICPSR). In short, a hands-on skill or a substantial knowledge shared in a lightening 10-minute presentation/talk in class. You can follow-up with online resources in discussion thread if you find it necessary (e.g. share the do-file used, or post the web-link to material you covered). This is a wide-open exercise so be creative, but remain relevant to our learning objectives for this course.

## Problem Sets

You are encouraged to work in study groups for your problem sets. However, you are required to submit your own problem sets. When you submit your problem sets, you must put names of the students with whom you have formed the study group. The best use of study groups is to discuss your doubts, learn from each other and debate the underlying concepts in each problem set. Problem sets are given with dual purpose: to *assess* how well you are doing in terms of your learning outcomes but more importantly, to give you an opportunity to *master* the concepts and skills you will be learning in this course.

## Term-Project

You will either conduct your own term-project or a team project (Maximum 2 students per team). Each team is expected to undertake a policy relevant research involving a statistical analysis of dataset of your choice. Only substantive requirement of the term-project is that your analysis should include at least one example of advanced statistical analysis you have learned in the class. Each team will find a dataset for the term project, make a fifteen minutes presentation towards the end of the semester and submit a scholarly paper.

The library data services guest lecturer will introduce the websites that contain good datasets. In addition, several data sources will be listed on course website. Please note that these are indicative sources; please feel free to select datasets from other sources.

The final output will be in a form of a manuscript that is concurrently submitted to a scientific journal in your field and a fifteen minutes conference style presentation (consider submitting your abstract to a real conference as well). Alternatively, you should provide a manuscript draft that is intended for publication and one page statement explaining the improvements needed before it could be submitted to a professional journal. In this one page statement, you should also include rough plan for completing the remaining work required to make it a publishable manuscript. You are also encouraged to submit your manuscript to a conference of your interest. The paper is expected to be approximately 20 pages (double spaced, regular margins, normal font size). Both the presentation and paper should cover: (i) background and motivation for working on your policy research; ii) short literature review focusing on statistical studies on your topic; iii) key research/policy question/s and hypothesis; (iv) data description and descriptive analysis of interest variables in the dataset; (v) the model specification; (vi) summary and policy implications of your findings and (vii) limitations and way forward.

The descriptive analysis should include tables and figures constructed using STATA. A good paper inadvertently refers and explains each table, plot and graph in the main text. The model section should specify the model used for the analysis.

The presentation is expected to follow a professional conference format. Each team will present within the fifteen minutes time limit (in both directions: not too short, not too long). As a thumb rule, 1 minute per PowerPoint slide (so 15 slide in total). If you are working in a team, both members must present. Even if each member prepares the slides he or she will be presenting, a professional presentation is always coherent, complete and transition is absolutely seamless and smooth. It always helps to practice the presentation a day in advance in front of your friends, family or mirror.

It is very important to cite all the papers that you review, source of the datasets and material used in the analysis (including software packages) – See UMB Policy below. Each team is expected to submit an electronic form of the presentation (ppt/x format). Each team is also expected to submit both a hardcopy and electronic forms (doc/x format) of the paper.

**Code of Conduct and Academic Integrity**

It is the expressed policy of the University that every aspect of academic life – not only formal coursework situations, but all relationships and interactions connected to the educational process – shall be conducted in an absolutely and uncompromisingly honest manner. The University presupposes that any submission of work for academic credit is the student’s own and is in compliance with University policies, including its policies on appropriate citation and plagiarism. These policies are spelled out in the Code of Student Conduct here: <https://www.umb.edu/life_on_campus/policies/community/code>. Students are required to adhere to the code of Student Conduct, including requirements for academic honesty, as delineated in the University of Massachusetts Graduate Catalogue and relevant program student handbook(s).

**Accommodations**

The University of Massachusetts Boston is committed to providing reasonable academic accommodations for all students with disabilities. This syllabus is available in alternate format upon request. If you have a disability and feel you will need accommodations in this course, please contact the Ross Center for Disability Services, Campus Center, Upper Level, Room 211 at 617.287.7430. http://www.umb.edu/academics/vpass/disability/

After registration with the Ross Center, a student can request accommodations at any time; we recommend that students inform the professor of the need for accommodations by the end of the Drop/Add period to ensure that accommodations are available for the entirety of the course.

**Detailed Schedule**

**Week 1: Jan 27**

Interest and Topic Selection Survey

Introduction and Course Overview

Ordinary Least Square (OLS) Review

Gauss Markov (GM) Theorem and Classical Linear Model (CLM)

*Readings: Wooldridge Chapter 1, Chapter 3.1-3.5; 4.1, Suggested Review: Appendices A, B and C*

**Week 2: Feb 5**

Final list of topics announced

Carrying out an empirical project

Guest Lecture by Librarian from Healey Library on Finding Statistical Datasets

*Readings: Wooldridge Chapter 19*

**Week 3: Feb 10**

Interaction terms

Log-linear regressions

Categorical dependent variables

Logit (logistic) regression

Probit regression

*Readings: Woodridge Ch. 6-7, Ch. 17*

**Problem Set 1 Distributed**

**Week 4: Feb 17**

Poisson and Negative Binomial regressions

Hosmer & Lemeshow Test

Zero-inflated Poisson and Negative Binomial regressions

*Readings: Woodridge Chapter 17*

**Week 5: Feb 24**

Tobit regressions

Selection Bias:

Heckman’s Sample Selection Model

Propensity Score Matching

AIC & BIC

**Problem Set 1 Due**

*Readings: Wooldridge Chapter 17.2 and Wooldridge Chapter 17.*

**Week 6: Mar 2**

**Problem Set 1 Solution Review**

**Review of Topics covered so far**

**Project Proposal Presentations**

**Term-Project Proposals Due**

**Week 7: Mar 9**

Heteroscedasticity and Weighted Least Square (WLS)

Robust Regression

Factor Analysis

*Readings: Woodridge Chapter 8.1 to 8.4*

**Problem Set 2 Distributed**

**Spring Break March 13 – March 20**

**Week 8: Mar 23**

Instrumental Variable (IV) / 2 Stage Least Squares (2SLS)

Path Analysis

Structural Equation Modeling

*Readings: Woodridge Chapter 15.1 to 15.5*

**Problem Set 2 Due**

**Week 9: Mar 30**

**Problem Set 2 Review**

Simultaneous Equations Models

Pooled sample analysis

Stata Session for Panel Data Manipulations (Missing Data)

*Readings: Woodridge Chapter 16.1 to 16.3, 13.1 to 13.4*

**Week 10: Apr 6**

Panel Data Analysis: Fixed and Random Effects Estimations & Hausman Test

Autocorrelation, Finite Distributed Lag (FDL) and Generalized Least Square (GLS) models

*Readings: Woodridge Chapter 10, 12 and 14*

**Problem Set 3 Distributed**

**Week 11: Apr 13**

Nonlinear regressions with panel data

Hierarchical Linear Model (HLM)

*Readings: TBA*

**Week 12: Apr 20**

Difference-in-Differences

Lagged Variables

(Regression Discontinuity Design)

**Problem Set 3 Due**

*Readings: TBA*

**Week 13: Apr 27**

**Problem Set 3 Review**

Simulation Methods:

Discrete Event Simulations

Agent-based Modeling

Micro-simulations

(Dynamic Programming)

*Readings: TBA*

**Week 14: May 4**

Spatial Analysis Methods:

Point Pattern Analysis

Location Allocation Modeling

Spatial Regressions

(Non-parametric data analysis)

*Readings: TBA*

**Week 15: May 11**

**Term-Project Presentations**

**Week 16: May 18 (Class does not meet)**

**Term Paper Due**