PADM-GP 4502
Using Large Datasets in Policy Research
Spring 2021

Instructor Information

- Professor Kristina Arakelyan
    - Email: kristina.arakelyan@nyu.edu
    - Office Address: Zoom
    - Office Hours: Mondays, 7pm – 8pm

Lectures:

- Dates: 3/24/2021 - 5/5/2021
- Day: Wednesdays
- Location: Zoom
- Time: 6:45pm – 8:25pm

Course Description

This half-semester course will focus on the analysis of data. We will discuss cleaning raw data – including trimming, variable transformations, and dealing with missing data – before turning to complex survey data. We will discuss how regression analysis differs when using complex survey data. Students will take real data and produce a cleaned version, as well as perform simple analyses using multiple regression. One key skill you will learn in this class is Stata, a commonly used statistics package. I will teach the basics, but if you have not used it before, you will likely need to spend a bit of extra time becoming acquainted with the program.

Course Prerequisites

CORE-GP 1011 is the prerequisite for the course. We will be doing lots of coding work using Stata and you will be expected to know the basics of regression analysis. As such, an advanced statistics course at Wagner, like PADM-GP 2902, is recommended. If you do not have experience with Stata, this class will require a good bit of extra work to get up to speed.

Textbooks

There are no required textbooks. However, there are some recommendations, especially if you will continue with data analysis in the future:

- Acock, A. C. (2008). *A gentle introduction to Stata*. Stata press. College Station, TX: Stata Press. [This is a good guide for students looking to build a solid foundation in using Stata]
- Long, J. S., & Long, J. S. (2009). *The workflow of data analysis using Stata.* College Station, TX: Stata Press.
- Mitchell, M. N. (2008). *A visual guide to Stata graphics.* College Station, TX: Stata press.
- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata.* College Station, TX: Stata press.

Software

We will use Stata in this class (pretty much any version is fine, though I use Stata 16). You should purchase a student license, either temporary or permanent. On the Stata website (https://www.stata.com/order/new/edu/gradplans/student-pricing/). Stata/IC will be sufficient for this class. You can purchase a perpetual license for $225, an annual license for $94, or a six-month license for $48. As an alternative, you can access Stata through NYU's Virtual Computer Lab (https://www.nyu.edu/life/information-technology/instructional-technology-support/instructional-technology-tools-and-services/virtual-computer-lab.html). This option is free, but it requires internet connectivity whenever you wish to use Stata. In addition, students have reported connectivity issues in the past.

If at all possible, you may wish to have two monitors to make class easier. I will be using Stata and sharing my window through Zoom, but you will also want to be able to use Stata on your own computer. Alternatively, you might consider using a second computer/tablet to log into the Zoom meeting.

NYU Classes Site

I will post assignments and datasets on our classes site. You will turn in all assignments through the site, as well.

Academic Integrity

Academic integrity is a vital component of Wagner and NYU. All students enrolled in this class are required to read and abide by Wagner's Academic Code. All Wagner students have already read and signed the Wagner Academic Oath. Plagiarism of any form will not be tolerated and students in this class are expected to report violations to me. If any student in this class is unsure about what is expected of you and how to abide by the academic code, you should consult with me.

Henry and Lucy Moses Center for Students with Disabilities at NYU

Academic accommodations are available for students with disabilities. Please visit the [Moses Center for Students with Disabilities (CSD) website](#) and click on the Reasonable Accommodations and How to Register tab or call or email CSD at (212-998-4980 or [mosescsd@nyu.edu](#)) for information. Students who are requesting academic accommodations are strongly advised to reach out to the Moses Center as early as possible in the semester for assistance.

Course Requirements and Grades

Course grades are based on the following:
- Homework (five, 10% each – lowest one dropped)
- Final assignment (40%) – Please note that there is no final exam for this course.
- Participation (20%)

Late Policy

For both the homework and final, barring extenuating circumstances, you will lose 5% for every 24 hours that the assignment is late (e.g., If you earn a 95/100 on your homework assignment but turn it in one day late, you will receive a 90/100). The absolutely last day to turn in assignments is Saturday, May 15th because grades are due soon after.

Participation

Barring technical difficulties, you should have your camera on during class. If you need to miss a class, please try to let me know. Each class builds on skills covered in the previous class; so, you should make sure to review missed material/get caught up not to fall behind. Excessive unexcused absences from class may result in a grade reduction.

NYU's Calendar Policy on Religious Holidays

[NYU's Calendar Policy on Religious Holidays](#) states that members of any religious group may, without penalty, absent themselves from classes when required in compliance with their religious obligations. Please notify me in advance of religious holidays that might coincide with class.

If you have questions:

Please e-mail me at kristina.arakelyan@nyu.edu or pop by office hours (Mondays, 7-8pm via Zoom). I try to respond to e-mails within 24-48 hours during weekdays.

Disclaimer: Elements of this syllabus are subject to change by the instructor. Students will be given sufficient notice of any changes made in the service of the course.

Course topics

Class 1 (3/24) – Introduction

Topics:

1. Introduction
    - Syllabus
    - Expectations
2. Use of data in policy (an introduction to and basic definition of datasets, bigdata, administrative data, etc.)
3. Data workflow (basic steps from data acquisition to finished project)
    - Importance of documenting all steps in a project
4. Types of variables
5. Formulating research questions

Class 2 (3/31) – Stata

Topics:

1. Using do-files to document work
    - Importance of comments in do-file
2. Using log files
3. Measures of central tendency and dispersion
4. Computing descriptive statistics and visuals (graphs) in Stata
5. Standardization
6. Bivariate analysis

Due: ***Homework 1 (Due by 11:59pm on Tuesday, April 6th)***

Class 3 (4/7) – Data Prep I

Topics:

1. Visualization as a first step
2. Cleaning
    - Trimming, winsorizing
    - Generating new variables and replacing old ones
3. Labeling variables and values

Due: ***Homework 2 (Due by 11:59pm on Tuesday, April 13th)***

Class 4 (4/14) – Data Prep II

Topics:

1. Missing data in Stata

2. Dealing with missing data
    a. Skip sequence?
    b. Dropping missing data
    c. Imputing missing data

Due: ***Homework 3 (Due by 11:59pm on Tuesday, April 20th)***

Class 5 (4/21) – Data Prep III

Topics:

1. Variable transformations and interpretations
    a. Logs
    b. Quadratics
2. Merging
    a. Data modules

Due: ***Homework 4 (Due by 11:59pm on Tuesday, April 27th)***

Class 6 (4/28) – Working with Complex Survey Data I

Topics:

1. Complex survey design
    a. Stratification
    b. Clustering
    c. Weighting

Due: ***Homework 5 (Due by 11:59pm on Tuesday, May 4th)***

Class 7 (5/5) – Working with Complex Survey Data II

Topics:

1. Dealing with complex survey data in Stata
    a. *svyset* command
    b. Implications of survey design for regression output
2. Becoming a consumer of quantitative research in public policy
3. Practical issues with sampling

The final assignment is by 11:59pm on Thursday, May 13th