NYU | ROBERT F. WAGNER GRADUATE SCHOOL OF PUBLIC SERVICE

PADM-GP 4502
Using Large Datasets in Policy Research
Fall 2024

Instructor Information

- Professor Kristina Arakelyan
  - Email: kristina.arakelyan@nyu.edu
  - Office Hours: by appointment

Lectures:

- Dates: 9/9/2024 – 10/21/2024
- Day: Mondays
- Location: 70 Washington Sq S (Bobst), Room LL141
- Time: 4:55pm – 6:35pm

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

PADM-GP 2902 is the prerequisite for the course.

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. (2018). *A gentle introduction to Stata* (6th Edition). 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. (2021). *A visual guide to Stata graphics* (4th Edition). College Station, TX: Stata press.
- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata* (2nd Edition). College Station, TX: Stata press.

Software

We will use Stata in this class (pretty much any version is fine, though I use Stata 18). You should purchase a student license, either temporary or permanent at https://www.stata.com/order/new/edu/gradplans/student-pricing/. Stata/BE will be sufficient for this class. You can purchase a perpetual license for $225 (if you go with this option, you might consider getting Stata/SE for $425 to be able to work with bigger datasets), an annual license for $94, or a six-month license for $48. As an alternative, you can access Stata through NYU IT Student Technology Center and Virtual Computer Lab (https://www.nyu.edu/life/information-technology/computing-support/software/software/stata.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.

NYU Brightspace

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. The use of generative AI tools is strictly prohibited in this course; any use of such tools will be considered a violation of the Wagner Academic Code. 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.

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.

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

Please let me know in advance if you are having difficulty completing an assignment on time so that we can discuss your situation. Assignments that are late without notification will automatically be downgraded 5% per day. The last day to turn in assignments late, with penalties, is October 27th.

Participation

Since we only meet once a week, I expect everyone to attend every class. If for any reason you will be unable to attend class, please let me know in advance.

If you are feeling unwell, please rest and focus on your health. Absences on account of illness, quarantine, travel disruptions, and other extenuating circumstances will, of course, be excused.

Syllabus Modification

I reserve the right to modify the syllabus based on input, class pace, and other factors as the term progresses.

Course topics

Class 1 (9/9) – 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. Finding data
4. Formulating research questions
5. Data workflow (basic steps from data acquisition to finished project)
    - Importance of documenting all steps in a project
6. Feedback loop

Class 2 (9/16) – 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
7. Data challenge: educational data

Due: ***Homework 1 (Due by 11:59pm on Sunday, September 22nd)***

Class 3 (9/23) – 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
4. Data challenge: healthcare data

Due: ***Homework 2 (Due by 11:59pm on Sunday, September 29tht)***

Class 4 (9/30) – 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
3. Data challenge: social service data

Due: ***Homework 3 (Due by 11:59pm on Sunday, October 6th)***

Class 5 (10/7) – Data Prep III

Topics:

1. Variable transformations and interpretations
    a. Logs
    b. Quadratics
2. Merging
    a. Data modules
3. Data challenge: criminal justice data

Due: ***Homework 4 (Due by 11:59pm on Sunday, October 13th)***

Class 6 (TUESDAY 10/15) – Working with Complex Survey Data I

Topics:

1. Complex survey design
    a. Stratification
    b. Clustering
    c. Weighting
2. Data challenge: environmental data

Due: ***Homework 5 (Due by 11:59pm on Monday, October 21s)***

Class 7 (10/21) – 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. Practical issues with sampling
3. Data ethics
4. Being a consumer of quantitative research in public policy

The final assignment is due by 11:59pm on Sunday, October 27th