PA 5932 Working with data: finding, managing, and using data

1.5 credits
Fall 2022, Mon/Wed 11:15 AM – 12:30 PM @ HHH 85

Instructor

Yi Wang (wang8262@umn.edu)
Office Hours: Mon 1:00 – 3:00 PM. HHH Rm 55, Cube 5519.
If the office hours do not work for you, email me and we can figure out a time (I normally respond to emails within 24 hours).

Course Description

This course will teach you how to find, download, manipulate, combine, and organize quantitative data to answer policy questions in efficient ways. You will gain hands-on experience working in Excel and Stata through class exercises and assignments. You will also be able to apply the skills learned to your own policy analysis. The software applications used in this class include Excel, Stata, and Adobe Illustrator.

You will learn best if you have a particular data set in mind that you would like to explore through this class. However, not having a data set in mind is also fine! I will provide you with data sets to work with.

After successful completion of this course, you should be able to:

1. Determine where to find data for policy-related topics.
2. Clean and manipulate 2nd hand data sets to answer policy-related questions.
3. Automate table-making for descriptive statistics and regression results.
4. Produce publication-quality figures that help you to answer policy-related questions.
5. Manage your data effectively and ensure reproducibility of your policy analysis.

A typical class will start with an introduction to the topic and then be followed by class exercises.

This class is NOT meant to teach you all the finer details of Excel and Stata. The main purpose is to introduce you to tricky obstacles to recognize and common pitfalls to avoid when using data in your own work. Data (like life) is messy and the practical experiences from the assignments offer a variety of common policy questions and a brief introduction to different online data resources for finding and working with data in Excel and Stata (mostly Stata).

Texts and Technology

There is no textbook for the course. Stata and Adobe Illustrator are available on all computers in Humphrey Lab 85. You can use lab computers during non-class times. Lab schedule is posted outside the lab door for your reference. You can also access lab computers remotely via the University of Minnesota's AppsToGo. To use the AppsToGo service on your own device, follow the instruction here.
Other useful resources for working with Stata:

- Stata Support - Accessible by typing "help" into the Stata command line.
- Statalist
- UCLA's statistical consulting office
- UWM's social science computing cooperative
- Guttmacher Institute Coding Style Guide
- Statistics help

Evaluation

The course grade will be based on a set of homework assignments and a final project. Each homework assignment should lead to your final project. There will be 6 homework assignments worth 80 points in total. The final project will be 20 points. The homework assignments will always be handed out on Wednesdays and will be due the following Wednesday. All assignments will be submitted individually. Late homework will receive a 10% deduction from the possible points for the assignment.

Assignments: 80%
Final project: 20%

Assignment Options

Considering that everyone is coming with a different Stata/Excel background with different goals to achieve through this class, there will be two options that you can choose for working on your homework assignments and your final project. This class will be most useful if you get hands-on experience working on a data set that you would like to explore. Therefore, I highly recommend going for option B if you would like to get the most out of this class during the short 7-weeks course period.

Option A: You will work on a data set that I provide. Your homework assignments and final project will be based on completing the task that I ask you to do.
Option B: You will work on a data set that you would like to explore. You will work on the basic skills we learn in class and depending on your interest, you will have the opportunity to learn beyond what is listed on the course syllabus. You and I will come up with a final goal of the project that you would like to accomplish.

Grading

93% and above A
90% and above A-
87% and above B+
83% and above B
73% and above B-
77% and above C+
73% and above C
70% and above C-
67% and above D+

Course Schedule

Timeline	*Focus*	*To Do*
Week 1	**Oct 26 - Introduction and finding data** • Syllabus • Where to find data	Meet with the instructor and discuss the assignment options.
Week 2	**Oct 31 - Data analysis in Excel (1)** • Excel Basics **Nov 2 - Data analysis in Excel (2)** • Make Pivot Table your best Friend	Homework 1: Due 9am Wednesday November 03
Week 3	**Nov 7 - Introduction to Stata (1)** • Do-files and Logs • Use "help" • Initial examination of the data • Recode & Create variables • Weight • Two-way tables / Crosstabulation **Nov 9 - Introduction to Stata (2)**	Homework 2: Due 9am Wednesday November 10
Week 4	**Nov 14 - Introduction to Stata (3)** • Collapse • Reshape • Scatter plot **Nov 16 - Introduction to Stata (4)**	Homework 3: Due 9am Wednesday November 17

Week 5	**Nov 21 - Introduction to Stata (5)** • Workflow • Merge and Append • Automate table-making (descriptive statistics tables & regression tables) **Nov 23 - Data visualization** • Use Adobe Illustrator to make your graphs look better	Homework 4: Due 9am Wednesday November 24
Week 6	**Nov 28 - Data encryption** • Encrypt a variable **Nov 30 - Stata programming (1)** • Macros and Loops	Homework 5: Due 9am Wednesday December 01
Week 7	**Dec 5 - Stata programming (2)** **Dec 7 - Bringing it all together** • Answer coding questions	Homework 6: Due 9am Wednesday December 08
Final Project: Due 9am Wednesday December 14 Applying all the skills in your final project that uses messy data to produce tables and graphs in Stata.		

Acknowledgement

Teaching is a collective endeavor, and this class has particularly benefited from generous sharing by Tom VanHeuvelen, Kenn Chua, Sangyoo Lee, and Anna Bolgrien.