Homework 1: STATA Practice

Prof. Tzu-Ting Yang

March 21, 2023

Instruction

- You should submit the Homework 1 before 4/17 class (3:30pm, 4/17).
- There is no late submission (you get zero point if you submit Homework 1 after the above deadline).
- The Howework 1 will account for 5 points in your final grade.
- Total grade of Howework 1 is 100. Question 1 accounts for 20, Question 2 accounts for 20, Question 3 accounts for 20, Question 4 accounts for 20, and Question 5 accounts for 20.
- This homework will help you practice STATA commands and let you start to work on your term paper.
- Your answers should include the STATA commands you use, explain their meanings, and your findings.
- Please upload your **answer sheet** and **log file** using this link: **https://reurl.cc/2WjY5X**.
- Format of the file name: StudentID_YourName. For example, r95323010_Tzu_Ting_Yang

1 Read Data

1. Briefly introduce your dataset.
2. If your dataset are stored in STATA format (.dta), you can utilize **use** command to load dataset. If not, utilize **import**, **infix**, or other commands to transform your data into STATA.

2 Examine Data

1. Pick any variable and use **sum** to get the median and mean of that variable.
2. Pick any variable and use **tab** to check the frequency of that variable.
3. Any variable has missing value? Utilize **inspect** or **codebook** to check them.
4. Utilize **duplicates** to examine whether your dataset has exactly the same observation in terms of the values of all variables.
5. Utilize **duplicates** to check whether the specified variable (pick any variable you want) can uniquely identify the observations.

3 Create Sample for Analysis: Part 1 & Part 2

1. Utilize **generate** to create a variable that you are interested in (e.g. treatment or outcome variables in your term paper) and breifly explain the meaning of that variable.
2. Utilize **egen** to create a variable that might be useful for your data analysis and breifly explain the meaning of that variable.
3. Utilize **label** to define value label and assign value label to a variable in your dataset (pick any variable you want).
4. Pick any variable and use **recode** to change the values of that variable.
5. Use **collapse**, **append**, or **merge** to create a new dataset that is useful for your analysis. Briefly describe this new dataset.

4 Visualize Data

1. Use **histogram** to display a variable that you are interested in. Please make sure readers can understand your graph (Hint: use option **xtitle**, **ytitle**, **xlabel**, and **ylabel**).
2. Use **graph twoway** to display the relationship of any two variables that you are interested in. Briefly explain your findings.

5 Prelimilary Analysis

1. Use **regress** to investigate the relationship between your treatment variable and outcome variable (without controlling any covariates). What is your finding?
2. Do you think the above estimate is causal effect of treatment? Is there any omitted variable bias (OVB, selection bias)? Use the **OVB formula** to discuss how the omitted variables bias your causal estimate in the regression.

3. Include relevent covariates into your regression and report your estimated coefficient on treatment variable. Discuss why the covariates you control are not bad controls?