Assignment 1

ECN 620 - Applied Economic Analysis – Fall 2024

Instructions:

1. The assignment is worth 10% of the final grade.
2. Assignment is due on **October 9th at 7 PM**
3. No late assignments will be accepted.
4. Students may work in groups of 4 individuals or less.
5. **But each student must write and submit their own do_file**.
6. Group members' names must be listed in the do_file. If you fail to list your collaborators, you will receive a score of zero.
7. There are two parts to the assignment. Part 1 is worth 80 marks. Part 2 is worth 20 marks.
8. Within each part, all subquestions are equally weighted.
9. You are required to enter your code and responses in the do-file provided on Brightspace.
10. Please remember to save your do-file using the following naming convention: **A1_YourStudentNumber_YourName**
11. Before doing the assignment, you will first need to get access to Stata. You can have access to Stata through the "Arts Virtual Desktop". Detailed instructions are available on the course webpage.
12. In cases where applicable, you will receive full marks for a question if I can run your code and obtain the answer you provided. If you provide the correct answer but the code does not work, you will receive half the points.

There are two datasets for this assignment:

- The file **A1_Data1.dta** contains information on monthly earnings (wage), years of education (educ), age (age), experience (exp), marital status (married) and race (race) for a sample of **married** male workers in the US. Variable married is 1 if married and 0 otherwise. Variable race is 1 if white and 0 if non-white

- The file **A1_Data2.dta** contains the same information but for a sample of **single** male workers in the US.

Download the datasets and do-file posted on Brightspace and complete the steps below.

Part 1: Basic Data Management (each question is worth 8 points)

1. Load the dataset **A1_Data1.dta** and add observations from the **A1_Data2.dta dataset** (Hint: use *append*). Export and save the combined dataset as a CSV file: A1_merge_data.csv (use comma-separated format)

2. Load the file **A1_merge_data.csv** created in the previous question in stata and write a code to answer Q3-Q12 below.

3. Clean the sample by removing observations with missing information in the variable **wage**. Generate a variable indicating the unique id for each worker. Name this variable **worker_id** and add the following label **"unique worker id".**

4. How many observations are in the sample? What is the mean wage in the sample? What is the minimum wage? (2 digits)

5. Wages and earnings are often studied by first taking logarithms. Transform the wage variable by taking its logarithm, and report the mean value of the log(wage) variable.

6. How many workers have exactly a high-school degree (12 years of education)? How many workers have at least a college degree (16 years or more)?

7. How many non-white workers are in the sample? What is the average wage for those workers?

8. What is the highest wage in the sample? What are the race, marital status, and age of the highest-wage worker in the sample?

9. Construct a scatter plot figure showing the relationship between **log wage** and **education** level. Add the following options to your graph:
    - graph title: Relationship between wage and education
    - X-axis title: Wage
    - Y-axis title: Years of education

10. Test the hypothesis that white and non-white workers earn the same wage on average. Interpret in one or two sentences the results of your test.

Part 2: Randomization (each question is worth 10 points)

11. Set the random number seed to the number that corresponds to the sample size computed in Q4 and Generate a variable "x" with random numbers distributed

uniformly between zero and one (Hint: use runiform). Then, create a variable treatment equal 1 if "x" is above the mean value of "x".

12. Define a label for the values of the variable treatment such that: 0 represents untreated units and 1 represents treated units. Test the hypothesis that treated and untreated workers earn the same wage on average. Based on the variable Did the randomization in Q11 work? Why? (explain in one or two sentences)