**Econ 303, T1 of 2024, Assignment 1 [worth 10%]**

**Background**

To find out if receiving a scholarship would encourage more students to pursue a postgraduate degree, Professor Stuff has run a survey experiment with some of her students.

The survey consists of some demographic questions and two questions about students’ likelihood of enrolling in a postgraduate degree in the next 2 years. The two questions are identical except for

* one question asks students to consider their current financial situation
* one question asks students to imagine they would receive a $1000/month scholarship

Professor Stuff wants to use the difference in the answers between these two questions as estimate of the effect of receiving a scholarship on the likelihood of pursuing a postgraduate degree.

It is your job to produce some summary statistics of the students who participated in the experiment and analyse the results of this experiment.

**You are given the following information:**

1. The survey <https://vuw.qualtrics.com/jfe/form/SV_7OkChdMAX7jk7Uq>
2. The text of the survey (available on Canvas)
3. The Excel file with the survey responses (available on Canvas)
4. The Stata file **demographic.dta** (available on Canvas) which contains demographic information on all students who participated in the experiment as well as other students. This file contains the following variables

* Id = student ID necessary to merge demographic data with the survey data
* gpa = students’ grade point average
* Firstyear = a dummy variable which is 1 if the student is a first year student and 0 if the student is a 2+ years student.

**Your tasks**

1. Use Stata to produce a neat-looking summary statistics table of the students who participated in the experiment.

Hint: you need to merge the survey data and the demographic data. For a neat-looking summary statistics table, you might need to create sensible variable names and/or labels.

1. Produce a neat-looking results table which contains results from
   1. A regression of students’ stated likelihood of pursuing a postgraduate degree on a dummy which is 1 if students answered the scholarship question and 0 if the students answered the current financial situation question.
   2. The same regression from a) for the sample of 1st year students
   3. The same regression from a) for the sample of 2+ year students

Hints: to run this regression, you need to reshape the survey data. You should cluster at the student ID level. You don’t need to include any other control variables.

**You should hand-in**

1. One do-file (named “Lastname\_data\_creation.do”) that creates

* the dataset for the summary statistics
* the summary statistics table
* the dataset for the results
* the results table

1. The dataset for the summary statistics (named “Lastname\_sumstats.dta”)
2. The dataset for the results (named “Lastname\_results.dta”)
3. One pdf-file showing the summary statistics table and results table (named “Lastname\_tables.pdf”). These tables should be identical to the ones that are produced with the do-file.

More details on 1. (do-file)

* The do-file should be well commented and organized.

More details on 2. and 3. (datasets)

* All variables should have an intuitive name (e.g. age instead of Q2)
* all variables should have numerical values whenever sensible (e.g. numbers for different categories of completed education)
* all variables that have numerical values that are not intuitive should include a value label (e.g. education = 1 “high school or less”)
* all non-sensical values (e.g. age <18) should be set to missing

**Marking criteria**

Your assignment will be graded taking into account the following criteria

* Does the do-file produce one summary statistics table and one results table?
* Is the summary statistics table in the pdf identical to the one produced by the do-file?
* Is the summary statistics table sensible and good-looking? (e.g. only relevant information, logical order of variables)
* Is the results table in the pdf identical to the one produced by the do-file?
* Is the results table sensible and good-looking? (e.g. correct estimation, correct clustering, helpful text in table)
* Is the do-file understandable and well commented?
* Does the do-file produce the sumsats dataset?
* Is the sumstats dataset “clean” and intuitive? (e.g. no unnecessary variables and observations, helpful variable names, helpful labels, value labels where necessary)
* Does the do-file produce the results dataset?
* Is the results dataset “clean” and intuitive? (e.g. no unnecessary variables and observations, helpful variable names, helpful labels, value labels where necessary)

**How to solve this assignment?**

To get started, carefully watch the video solutions for tutorials 2 and 3. Beyond this, you will most likely need to try a few things, use the Stata help function, and search the internet. This can be challenging.

This is an individual assignment. You are not allowed to collaborate with other students. For example, you are not allowed to look at your peers’ code, copy code from others (except the tutorial sample solutions), or ask someone else to solve the exercise for you.

**I take plagiarism seriously. I may compare your code with the code of other students. I might also ask you questions about your own code to make sure that you’ve written it.**

If you are unsure if something is allowed, please ask me.