Part 0 – Open Stata, and make your own do-file

- Start Stata through the start menu button
- In de white command window type `doedit` to start de do-file editor. Place the Stata screen on the left and the do file editor on the right such that you can easily switch between the two.
- Save the empty do-file as a new do-file under an applicable name such as `ectrcs_w1.do` in a directory that you want to use for this course, for example `H:/ectrcs`
- In the first two lines of the do-file type

```
cls                             //this clears the screen
clear all                       //this clears the memory
cd "H:/ectrcs"                  //this is your path
```

- If you want to use your own computer type instead:

```
cd "~/ectrcs"                   //this is your MAC path
cd "c:/ectrcs"                  //this is your PC path
```

- Press control-s to save the file, and control-d to run it.
- Then follow two **Rules**:
  1. *In all subsequent questions that you will do* in Part 1 and 2 (and the following Computer Lab sessions) first type the command - or series of commands - that you want to do in the command window. Then, when you are certain that the commands are correct (i.e. no syntax errors), copy-paste them in the do-file on the left, and save the do-file. You can now always run the do-file at a later time, *essential if you want to replicate your results*. You can also use the menus to find a certain command, and paste it in the do-file.
  2. For each question below, when appropriate, use commenting in the do-file to denote the question and give verbal answers using the `//` statement. For example

```
//Clear
cls
clear all

//goto project folder
cd "c:/ectrcs"

//Part 1
//Q1
use w1_data_experiment1.dta

//Q2
desc
/* xborrow is the outcome variable
 treat is 0,1 treatment indicator
 other variables are background characteristics */

//Q3
sum
drop if age<17
```

Comment

Part 1 - The effect of information about student loans on loan take-up

In the first part of this computer exercise we will replicate part of the results in Booij *et al*. (2012) that were also discussed in the lecture. The authors present the results of a field experiment where students in higher education are given information about student loans.

Figure 1 in the paper (see attachment to the lecture notes) reveals that students know relatively little about the loans that are available to them. Also, students that know more seem to borrow more often compared to students that are uninformed about the loans. If this is a true causal relationship, providing students with information about loans should lead them to borrow more often - an objective of the government. We could test this by randomly giving students information about loans (treatment group) and others not (control group), and see whether loan take-up indeed increases. This is exactly what is done in the Booij *et al*. (2012) paper.

The data provided contains information on treatment status (reception of information), a binary variable on subsequent borrowing (`xborrow`) and some background variables. You will use this data to replicate[1] some of the results in the paper and investigate whether giving students information has increased their borrowing.

1. Download the data `w1_data_experiment1.dta` from **Cv,** and put it into your `H:\ectrcs` folder. Now open the data by typing `use w1_data_experiment1.dta`**.** Next, type `ssc install estout, replace` to install an auxiliary package for obtaining nicely formatted tables.
2. Inform yourself about the variables by typing `desc`. Is it clear to you which variables are outcomes (Y), treatment status (X), and control variables ($\boldsymbol{W}$)?
3. Investigate the data by typing `sum`. Do you spot outliers? Delete the obvious outliers by typing `drop if age<17`.
4. The virtue of an experiment is that we can be certain that the variable of interest is not related to other factors. The experimenter should still, however, try to convince other scientists that the treatment was indeed randomized. One way to do so is to check whether background variables $\boldsymbol{W}$ are related to treatment. Check with an F-test whether the background variables significantly explain the treatment status. Interpret the result. Type `help test` to figure out how to do a test on multiple coefficients in Stata.
5. When we are convinced that the treatment is truly random, we can estimate the causal effect by a simple regression of the outcome variable on the treatment. Type `reg xborrow treat` and interpret the result. Should the government start an information campaign to stimulate borrowing?

[1] You will not get the identical results because you will only work with a linear specification of the variable SES instead of a full set of dummies.

6. To obtain correct standard errors we should add the `robust` option. Do so and discuss the difference with the previous answer. Store the equation by typing `est store ols1`.
7. We can include other background variables in the regression by adding them after the variable `treatment` in the regression command. Discuss the main differences with the first regression that did not include covariates. Some variables appear very significant. Do they have a causal interpretation? Store the equation by typing `est store ols2`
8. The point estimate on the treatment variable changes from .0017378 to .0037585 when including background variables. You can show this comparison in one table by typing `esttab ols1 ols2, se b(a2) keep(treatment)`. The difference will not be significant in most cases because the control variables are not related to the treatment. Verify that both point estimates lie within the range of the 95% confidence interval of the other variable (use q. 6 and 7). This is not an appropriate test, but if both estimates do not lie in each other's range, it is an indication that they may differ significantly.
9. Finally we are interested in whether students that are at the beginning of their study (first two years) respond more to the treatment than older students that are likely to have already formed financial habits. First `gen begin=(studydur<=24)`. Then perform a regression where you interact the treatment with this variable by typing `reg xborrow treat c.begin#c.treat begin age ethnminor ses studydur ra female actrack dr loanexp, robust`. Is there a significant difference in treatment response for student at the beginning or the end of their study? You also see that the main effect of `begin` is included, is this necessary?

Part 2 - The effect compulsory homework on exam grades

In the second part of this computer exercise we will consider (simulated) data concerning the relation between compulsory homework and test scores, similar to the example given in the lecture and in S&W p477[525]. The context is that of an econometrics course where students are randomly assigned to make compulsory homework or not. The outcome of interest is the grade on the exam. Some students follow the course as part of their major, while others took it as an elective (non-major). We will not use the `robust` option in this exercise.[2]

10. Download and open the data by typing `use w1_data_experiment2.dta, clear`. Inform yourself about the variables by typing `desc`. Is it clear to you which variables are outcomes (Y), treatment status (X), and control variables ($\boldsymbol{W}$)? *Hint*: Is there only one outcome variable?
11. Investigate the data by typing `sum`. Do you spot outliers?
12. Assuming students are randomly assigned we can simply regress `reg examscore homework` to determine the causal effect of homework on test scores. Do you find a

[2] The reason is that at q16 we are going to pool the two regressions using the `suest` command, which calculates robust standard errors automatically and cannot take estimated models that already have robust standard errors as arguments.

positive effect? Is the effect large in terms of standard deviation units? Store the equation by typing `est store ols3`.

13. You are skeptical that the assignment of homework was independent of major status. To investigate this you type `tab homework major, col nofreq`. Interpret this table. Support your conclusion also statistically with an *F*-test by running a regression of homework on background characteristics. What do you conclude regarding the assignment procedure?
14. Now regress `reg examscore homework major male` and interpret the coefficient on the homework variable. Can we give this coefficient a causal interpretation? And what about the other covariates? Store the equation by typing `est store ols4`, and subsequently display both model specifications nicely next to each other by typing `esttab ols3 ols4, se b(a2)`.
15. To get a sense if the difference between the two estimates differ significantly, check whether they lay in each other's CI.
16. The appropriate statistical test is to construct the t-statistic of the difference $t = (\hat{\beta}_1^4 - \hat{\beta}_1^3)/SE(\hat{\beta}_1^4 - \hat{\beta}_1^3)$ - where 3 and 4 denote the estimated model - and see if that is larger than 1.96. The difficulty is to obtain the standard error that appears in the denominator. We have $SE(\hat{\beta}_1^4 - \hat{\beta}_1^3) = \sqrt{SE(\hat{\beta}_1^4)^2 + SE(\hat{\beta}_1^3)^2 - 2Cov(\hat{\beta}_1^4, \hat{\beta}_1^3)}$, where the two standard errors can be obtained from the two tables. The covariance between the two point estimates is not given, however, so you cannot calculate the *t*-statistic yourself.

    Fortunately, Stata can do the appropriate test for you by first pooling the two regressions using `qui suest ols3 ols4`, and then typing `lincom [ols4_mean]homework-[ols3_mean]homework`. This table gives the estimated difference and *SE* which you can interpret as if it were "normal" regression output. What is your conclusion regarding the change in the coefficient of interest? From the output, can you work out what the covariance term must have been?
17. Finally, the data also contains the variable dropout. Include this variable into the previous regression. Do you trust the output from this regression? Is it wise to include another outcome variable in the regression? (p477[524])

Reference

Booij, A.S., Leuven, E. & Oosterbeek, H., 2012. The role of information in the take-up of student loans. *Economics of Education Review* 31(1), 33-44.