ECO 102: Topics in Economics
Ecole Polytechnique, Spring 2022
Professors: Geoffroy Barrows, Benoit Schmutz
Teaching Assistants: Arnault Chatelain, Maddalena Conte

TD2: Monte Carlo and OLS Regression

In this exercise, you will run what is called a Monte Carlo experiment. You will generate fake data using stata's random number generator. You will then analyze the data. You will compute some descriptive statistics. You will also estimate parameters using OLS and display the results. Please generate a pdf with all your results and do a very short description of the results. To compile the pdf, you should use a text editor like Latex.

A Very Brief Introduction to Latex

Latex is a software for document preparation. It is widely used in academia for it reduces the formatting task to a minimum and makes it easy to type mathematical formula (in fact, almost all your classes, exercise sheets and exams are typed on Latex).

There are two ways to work with Latex. Either you download a local Tex editor on your computer or you work on an online one[1]. During these TDs to make your Tex editing easier I will ask you to all create a free account on Overleaf with your Polytechnique mail address. Now go on Moodle and download the ECO102_TD2 zip file. Once done you can import it on Overleaf clicking on New Project -> Import a Project.

You should have two documents on the very left panel of your project. The first one mystyle.sty contains common useful packages to code in Latex. The second one TD2_questions.tex is your main document. It is the script of the pdf you are currently reading.

I will ask you to type your answers to today's exercise directly below the question in this script so that at the end of the TD we can export a pdf containing both questions and answers[2].

Exercise

Open a do file, you will save your Stata commands in it. Other answers and comments should appear in your pdf files.

1. In this part, you will generate some random data, export descriptive statistics and plot the data.

   (a) Set the seed in stata. When running monte carlos, it is good to set the seed. Results depend on the seed.

   (b) For 100 observations, generate x and ϵ. Draw x from a uniform distribution on the range (0,10). Draw ϵ from a normal distribution with mean 0 and variance 25.

   (c) Now build y according to the formula

   $$y_i = \alpha + \beta * x_i + +\epsilon_i$$

   with $\alpha = 10$ and $\beta = 2$

[1] For the pros and cons of each see: https://tex.stackexchange.com/questions/193810/online-latex-editor-or-local-latex-editor.

[2] For a good, much more detailed, introduction to Latex see: https://en.wikibooks.org/wiki/LaTeX.

(d) Compute descriptive statistics for y , x and ϵ (min, max, mean, standard deviation). Export as a table using the command sutex and compile into your pdf file.

Table 1: Summary statistics for y, x and ϵ

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
y	19.108	8.07	0.956	35.808
x	4.821	2.831	0.053	9.488
epsilon	-0.534	5.324	-9.654	17.107
N		100		

(e) Plot y against x in a scatter plot. Export the figure and compile into your pdf file.
The graph of y against x in a scatter plot is below

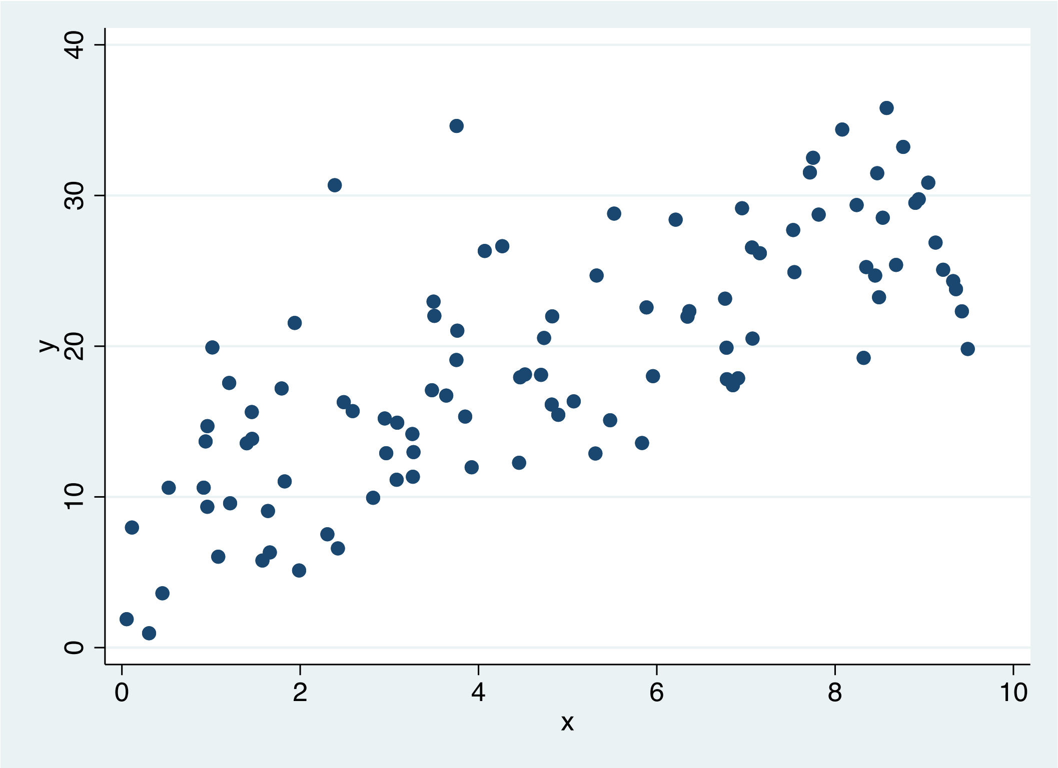

Figure 1: scatter graph of y against x

2. With your generated data, estimate α and β using OLS. Plot data along with OLS best-fit line. Include your estimated value of α and β on the graph (i.e., write the values somewhere on the graph). Export the figure and compile into your pdf file. Do you find that $\hat{\beta} = 2$? The data was generated using $\beta = 2$. Why don't you find $\hat{\beta} = 2$?

- Clearly, from the data we find from the OLS that $\beta = 2.147404$ which is not exactly equal to 2 because of the noise ϵ

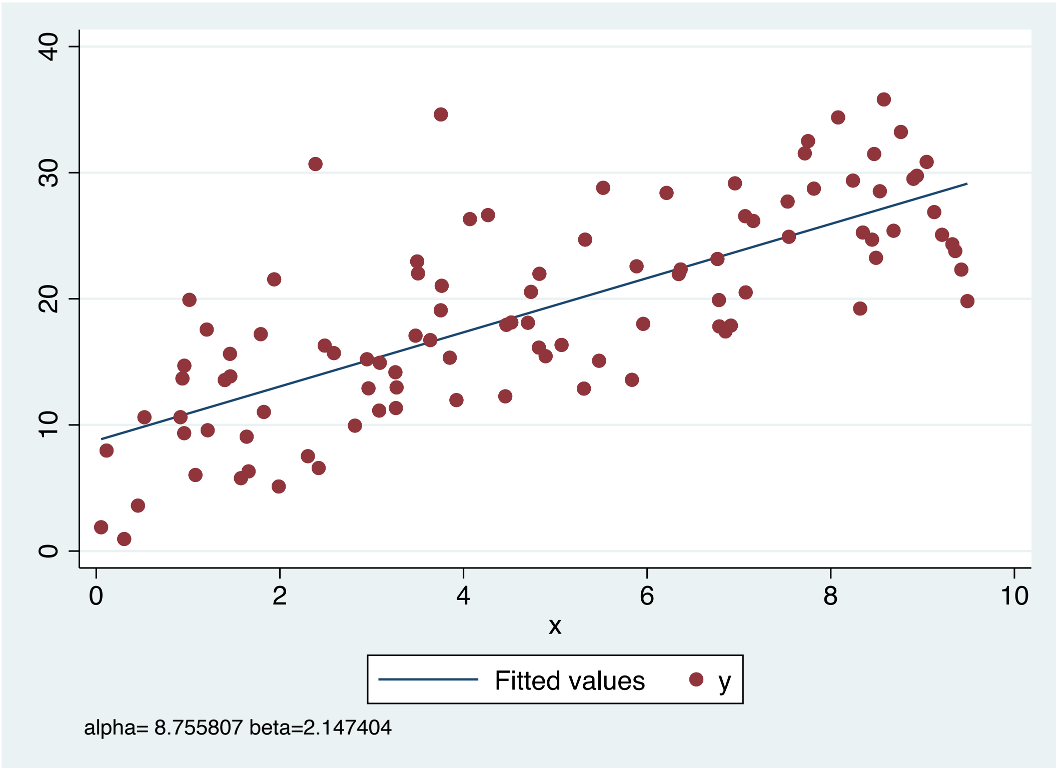

Figure 2: scatter graph of y against x - Note: $\hat{\alpha} = 8.755807$ $\hat{\beta} = 2.147404$

3. Repeat parts 1 and 2 for two replications of the data. I.e., randomly generate two different datasets, each with 100 observations. Plot the data along with the OLS best fit line and the estimated parameters in a graph separately for each replication. Export the figure and compile into your pdf file. Do you find the same $\hat{\beta}$ in the two samples? Why not?

   - From the OLS result of these two sample we can easily find that the $\hat{\beta}$ is not equal. This caused by the random sampling that the data of sample 1 and sample is not the same because of the *epsilon*. Therefore the $\hat{\beta}$ from the OLS result is not equal.

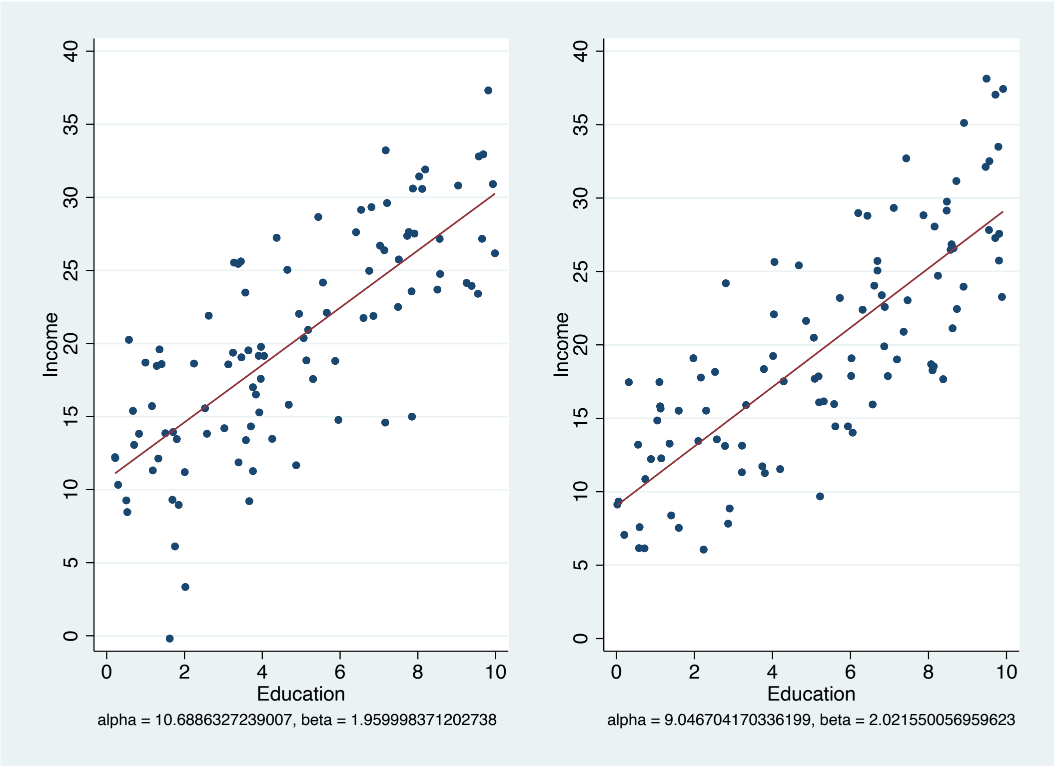

Figure 3: Comparison of two graphs

4. Now generate the data 1000 times. I.e., generate 1000 different datasets, each time randomly drawing x and ϵ and building $y_i = 10 + 2x_i + \epsilon_i$. For each replication, estimates α and β by OLS. Save each estimate $\hat{\beta}$ and the standard error of the estimate.

   (a) Plot the distribution of the 1000 $\hat{\beta}$'s. Report the mean and the median of the distribution on the graph. Export the figure and compile into your pdf file.

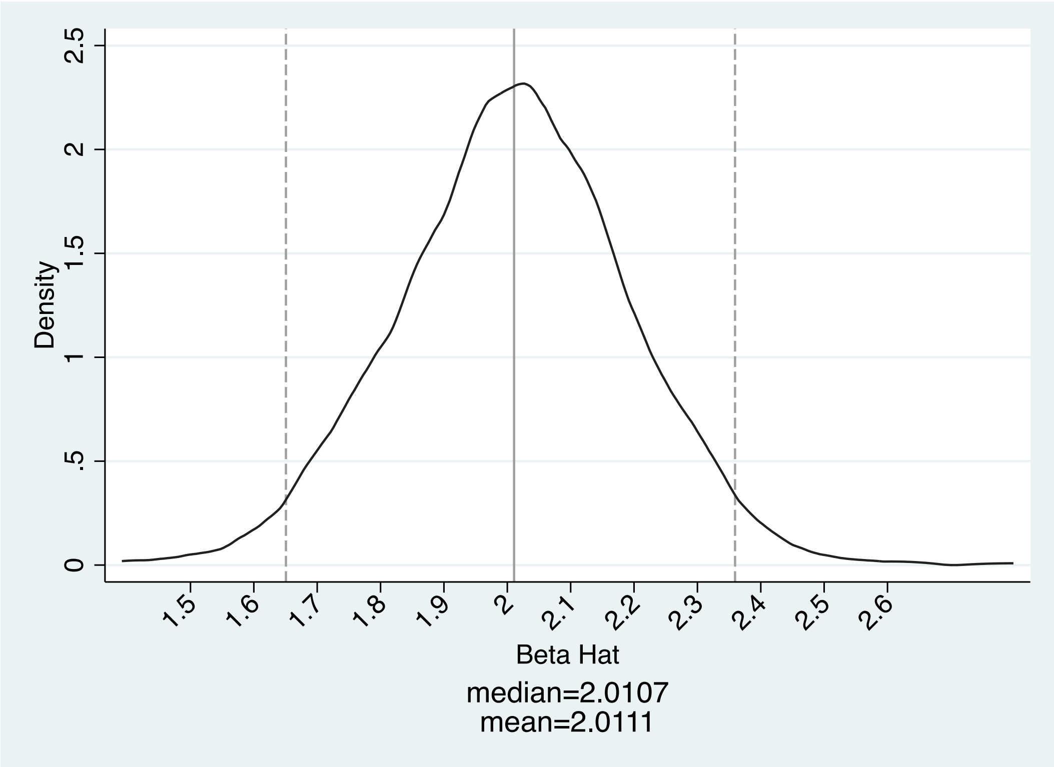

Figure 4: Graph of the Distribution of $\hat{\beta}$

(b) Compute the lower bound and upper bound of the 95% confidence interval for each of the 1000 $\hat{\beta}$'s. Plot the distributions of both the lower bound and upper bound. Report the share of replications for which the 95% confidence interval covers the true value of β. Export the figure and compile into your pdf file.

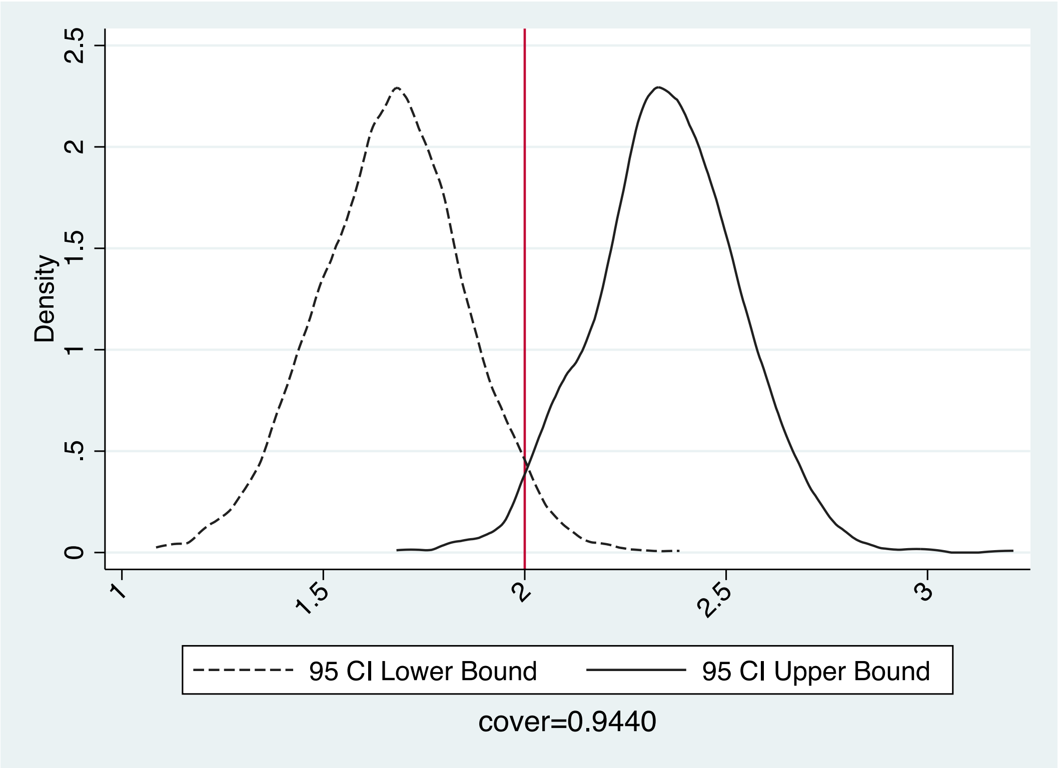

Figure 5: Graph of 95% Confidence Interval