Exam questionnaire

To be returned on AMeTICE by October, 2024

Pierre Michel

Description

In course example, we have studied the relationship between the life satisfaction and the GDP across European countries and over years. For this exam, you will have to perform a similar data analysis with two other variables X and Y. This project wil consists in four main tasks:

- Download data from Eurostat about variables X and Y. (Of course, do not choose life satisfaction and GDP).
- Import the data in R, describe the two datasets, clean them and join them together.
- Produce several descriptive statistics highlighted in graphs and tables. You will have to export these graphs and tables and include them in your document.
- Fit a first model where you regress Y on X. You will have to export the results into a table and include it into this document.

Remarks

- At the end, your document must be converted in pdf
- Do not include the R code
- Do not include the data
- You can include tables and figures

Description of the relationship of interest

1. Describe the two variables X and Y of interest.
2. Why is it interesting to study the relationship between X and Y?

Description of your data

3. How is measured X?
4. How is measured Y?
5. Which column(s) do you use to merge both datasets X and Y?
6. What is the unit of observation in your final dataset?
7. How many observations are there in your final dataset?

Descriptive statistics

8. What is the distribution of the variable X? Include an histogram and/or a density plot of X. Include it in the document.
9. What is the distribution of variable Y? Include an histogram and/or a density plot of X. Include it in the document.
10. Create a table of summary statistics with mean, standard deviation, minimum and maximum for both variables X and Y. Include it into this document.
11. Create a table of summary statistics where each row corresponds to one country and reports the name of the country, the mean of X (over the entire time period) and standard deviation and Y.
12. Plot the average evolution of X and Y across the years into two separated graphs. Include them into your document.
13. Select two countries and plot the average evolution of variables X and Y across the years for both in two separated graphs. Include these graphs to your document. Explain why you selected these two countries and why they provide important insights regarding the relationship between X and Y.
14. Compute change in X between first and last year for each country. Rank countries into a barplot in decreasing order. Do the same with variable Y. Include the two graphs into two separated panels of a same graph. Include it graph to your document.

Relationship between X and Y

15. Plot a scatterplot for the relationship between Y an X. Include a linear fit on the graph. Include this graph to your document.
16. Plot another scatterplot of the relationship between Y and X. Include a quadratic fit on the graph. Include it to your document.
17. Compute within-country changes over the year for variables X and Y. Plot a scatterplot between the within-country change in Y and the within-country change in X. Include a linear fit on the graph. Make Include it to your document.
18. Regress Y on X. Export the results into a table and include the table to this document.
19. Regress Y on X controlling for country fixed effects. Export the results of question 18 and 19 into a same table and include it to this document.
20. Regress Y on X controlling for year and country fixed effects. Export the results of question 18, 19 and 20 into a same table and include it to your document.