CSOC20010

Introduction to Computational Social Science II

UCD School of Sociology

Spring, 2021-2022

Module Coordinator: Assoc Prof Taha Yasseri

Week 4 Assignment: Scaling of GDP with Population in the EU

In this week's assignment, we are going to evaluate the scaling relationship between the GDP of cities and their population in EU cities. We know larger cities naturally produce more, but we would like to know if the relationship is linear, i.e. a city larger twice will produce exactly twice more, or it is super/sub linear meaning that the growth in GDP is faster/slower than a linear growth. To do that, take the following steps:

1. Download the data file "EU_Cities_Population_GDP"

2. Open the file in R, Excel, Python, or any other data analysis software you would like to use.

3. Transform all the data logarithmically, i.e., produce two new variables: Log(Population) and Log(GDP).

4. Fit a linear model to Log(GDP) as a function of Log(Population)

5. Create a scatter plot of the new variables (Log(population) in horizontal and Log(GDP) in vertical axes). Also, plot the line representing the linear model you fitted in step 4.

6. Create a second scatter plot that shows the original values (population and GDP) but uses a logarithmic scale on the axes.

7. Write a report that has a) the two figures you produced, b) the equation of the fit, and c) a few sentences of interpretation. Is the scaling linear, super-linear, or sub-linear? How can we explain this observation?