

In-class group assignment

Data Analysis 2 and Coding with R

18th of November, 2020

You need to carry out a complete single linear regression analysis, similar to our last class with regressing life expectancy on GDP per capita. Each of you are randomly assigned to a different groups (overall five groups) and you shall help each other in separate break-out rooms in zoom. The idea is it is an individual task, but you are encouraged to help each other. If you have any question or problem you should write to us (Misi or me) via slack and we will join and help in your breakout room.

What you need to do:

1. Download and clean the data.
2. Make a exploratory data analysis (EDA) on your key parameters.
3. Check and do transformations/scaling on your x or y variable with substantive and statistical arguments. (If you wish you can download a GDP measure or population, ect. and use them as a scaling variable to get ratios.)
4. Specify your models and run them with visualizations
5. Choose your final model and argue (substantive and statistical)
6. Do hypothesis testing on your coefficient(s) (e.g. equal to zero or not)
7. Analyse residuals - check the 5 worsts and bests
8. Create two graphs with: a) CI of regression line, b) prediction interval for the observations.
 - (a) Predict a country's outcome, where you do not observe the outcome in your data. Argue whether it is a feasible prediction or not.
9. If you are ready with the analysis create a presentation, using 'beamer' in Rmarkdown. This type automatically creates a pdf based presentation. You can check an example at the course's github repo

Please try to help each other during the process. If everybody is done in the group, you should start working together on a group presentation which present your main results. If we have time in the end of the class we are going to make a 5min/group presentations.

1 Groups, variables and dates

Last name	First name	Team	Y and X variables	Year
Fasih	Atif	A	Life Expectancy and Healthcare Expenditure	2000
Mariam	Bazzi	A	Life Expectancy and Healthcare Expenditure	2000
Maeva	Braeckeveldt	B	Saving and Income	2017
Pauline	Broussolle	A	Life Expectancy and Healthcare Expenditure	2000
Yuri	Cunha	B	Saving and Income	2017
Eszter	Diamant	E	CO2 and GDP	2000
Dominik	Gulácsy	E	CO2 and GDP	2000
Zsombor	Hegedus	D	Life Expectancy and Healthcare Expenditure	2017
Bruno	Helmechzy	C	CO2 and GDP	2015
Istvan	Janco	E	CO2 and GDP	2000
Li	Jia	C	CO2 and GDP	2015
John	Joyce	B	Saving and Income	2017
Ozan	Kaya	C	CO2 and GDP	2015
Fanni	Kiss	B	Saving and Income	2017
Abduvosid	Malikov	E	CO2 and GDP	2000
Viktoria	Mészáros	A	Life Expectancy and Healthcare Expenditure	2000
Attila	Serfőző	D	Life Expectancy and Healthcare Expenditure	2017
Tamas	Stahl	A	Life Expectancy and Healthcare Expenditure	2000
Kata	Süle	C	CO2 and GDP	2015
Terez	Szabó	D	Life Expectancy and Healthcare Expenditure	2017
Attila	Szűts	E	CO2 and GDP	2000
Karola	Takács	B	Saving and Income	2017
Cosmin Catalin	Ticu	D	Life Expectancy and Healthcare Expenditure	2017
David Gabor	Utassy	C	CO2 and GDP	2015
Xinqi	Wang	D	Life Expectancy and Healthcare Expenditure	2017

Table 1: Groups and names

You should use WDI library to get the data from Word Bank's database and for simplicity I recommend to use the following indicators:

Name	WDI code
CO2	EN.ATM.CO2E.PC
GDP for 2015	NY.GDP.PCAP.PP.KD
GDP for 2000	NY.GDP.PCAP.KD
Life Expectancy	SP.DYN.LE00.IN
Healthcare Expenditure	SH.XPD.CHEX.PP.CD
Saving	NY.GDS.TOTL.CD
Income	BN.GSR.FCTY.CD

Table 2: WDI code names