Professor: Héctor Bahamonde, PhD.
e:hibano@utu.fi
w:www.HectorBahamonde.com
Course: OLS.
TA: Valterri Pulkkinen.

INSTRUCTIONS

Using the randomly assigned dataset, and within the deadline established in the program, you will:

1. **Generate a research question**. An example might be *What explains Y?* To answer this question, use the *codebook*, and see which of the included variables might help you answer your research question. Also, remember to articulate your work based on the variables you do have.

2. **Generate a working hypothesis**. A working hypothesis is an answer (an assertion). An example might be *Y can be explained because of X_1*. In general, here you also explain *why* you think that's the case.

3. **Generate an alternative hypothesis**. Just like the working hypothesis, the working one is an assertion. The difference is that the alternative answers the question differently. For instance, here you can hypothesize that Y is rather explained by X_2 (not X_1).

4. **Hypothesis testing**. To do this you will need to perform at least one OLS model. Remember, Y should be continuous. Also, remember to use control variables. Always justify your control variables (tell us *why* you decided to include them). Maybe a (super) quick literature review and a short analysis of the problem might guide your decision-making process.

5. **Presentation**. *After testing both hypothesis, for which one do you find statistical support?* To answer this question, look at the (1) p-values, (2) standard errors, (3) r^2 of the model, and (4) effect sizes (β). Do *not* copy and paste R output into your presentation. Present a carefully designed regression table.

6. **Post-estimation**. Using a scatter plot, check (1) normality of your residuals/errors, and (2) homo/hetero-scedasticity. (a) Answer in your presentation: *Do your residuals behave according to the assumptions behind OLS theory?*

7. **In-class presentation ("conference")**. In the presentation you should talk about all these issues. Each presentation should last approximately 10-15 minutes. You will have to convince us that you did everything that's possible to find an appropriate model and that you tested all your assumptions. The rest of the class will not only listen to the presentation, but also will contribute to the discussion via constructive criticism.

8. **Products**:

    (a) **A script**: submit a R script with all your data analyses. Remember that the script should run properly and should not get stuck. For that, include in your code all the packages and libraries so I can replicate your work.

    (b) **One presentation**: a conference-style presentation with a Power Point.

DATASETS AND DEPENDENT VARIABLES

Library	Dataset	Codebook	Dependent Variable
library(AER)	data(Fatalities)	help(Fatalities)	Fatalities$fatal
	data(Guns)	help(Guns)	Guns$violent
	data(HousePrices)	help(HousePrices)	HousePrices$price
	data(Journals)	help(Journals)	Journals$price
	data(TeachingRatings)	help(TeachingRatings)	TeachingRatings$beauty