# Introduction to Statistics and Data Science

# Hand-in Assignment, Week 7 Hilary Term

Like your last assignment, you will need to create your own Jupyter Notebook (eg in Spyder or Colab) to carry out the analyses.

Remember you will need to include a code block at the top of your Jupyter notebook importing the relevant packages – you can copy this from one of my example notebooks in the online coursebook.

You will download the data as .csv files and load them from your local directory or your google drive if using Colab – see the online course book for a recap of how to do this if unsure!

Try to make your work readable for your tutor – include comments in all code. Comments are text preceded by hashtag #, explaining what you are doing, e.g.,

# find the mean of all columns in my dataframe

df.mean()

ASSIGNMENT

TASK: Write a short scientific report about people’s perceptions of their position in society and their happiness.

Use the data happy.csv as you did during the week7 tutorial. Remember that information about the data and variables is included in the tutorial section (15.8) of the online coursebook. This hand-in assignment should include the following sections, which follow the form of a scientific article.

1. Introduction: write a short introduction which includes a general description of the aims of the research. E.g., “In this research report, I will investigate…”
2. Data and methods: write a few sentences about the data, including which variables you are going to use. Mention how many missing data values there are. Here you should clearly describe the dependent variable and the explanatory variable, as well as the control variables you are using. Justify your choice of control variables.
3. Results: begin with some simple descriptive statistics (e.g., mean and standard deviation of the dependent variable). Include a regression table. You should include TWO regression models 1) an initial model, and 2) a more complex model that may include an interaction term if you choose. Write up the results in a paragraph. Mention statistical significance as well as the direction and the size of the effect. Report the R-squared and interpret.
4. Plot: Include one or two plots to illustrate your results visually.
5. Limitations: are there any limitations to your study? How good is your regression model? Mention here if any regression assumptions were violated. Do we know that the observed association is causal? Mention possible future research directions: e.g., what other variables might we need?
6. Conclusion: write two or three sentences summarising your study.

You will produce your report in the form of a Jupyter notebook. The notebook will contain text (following the format of sections outlined above) and code blocks (showing transparently how you processed the data).