

# INTL550: ADVANCED DATA ANALYSIS IN PYTHON

SERHAN TÜRKAN

## HOMEWORK 2

### Introduction

The main goal for many of us in life is to be happy. Happiness depends on many things such as economy, health, family, freedom, and trust. Although measuring happiness is empirically difficult, there are many studies.

### Hypothesis

Even if they say money does not bring happiness, happiness increases as the gross domestic product (GDP) increases in a country.

### Data Collection and Pre-processing

For this study, I have used World Happiness Report Dataset from Kaggle. The World Happiness Report is a landmark survey of the state of global happiness. The happiness scores and rankings use data from the Gallup World Poll. The columns following the happiness score estimate the extent to which each of six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity – contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world's lowest national averages for each of the six factors. The data collected was modified by selecting only the required columns and removing the unwanted data. Afterwards, the data was checked for invalid entries such as NaN values or null (empty) values. The invalid entries were removed by list-wise deletion. The test file written performs tests to validate the data input.

### Linear Regression Model

The mathematical calculation for the linear regression was derived as follows:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1},$$

$$\sigma^2 = \frac{e'e}{n - k - 1},$$

$$e = y - \hat{y},$$

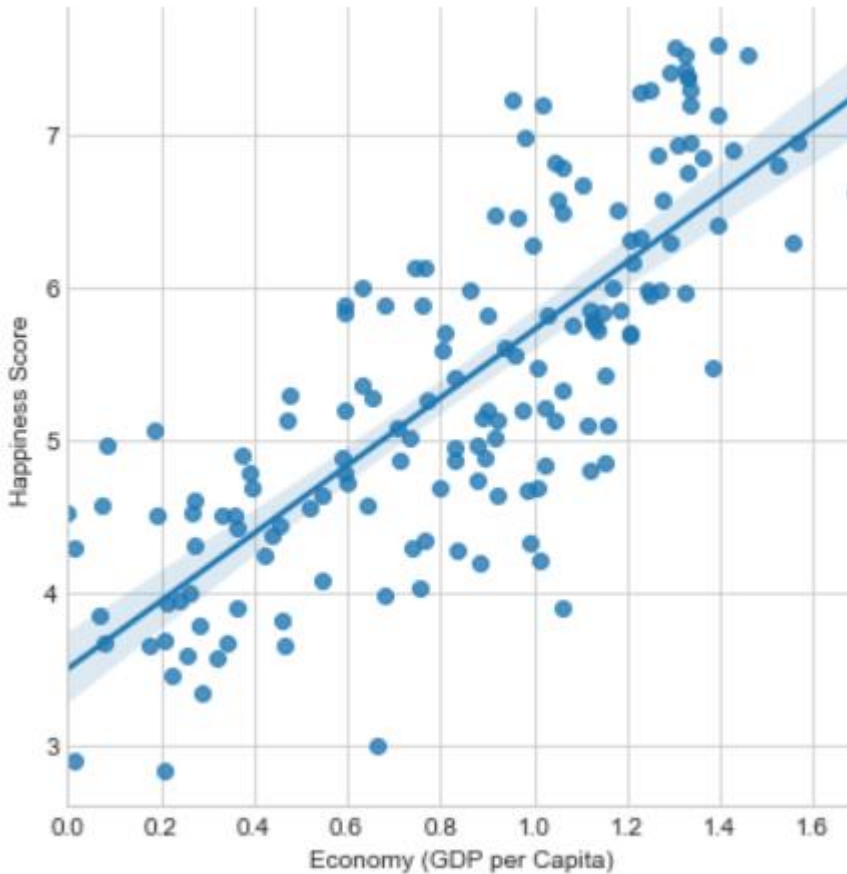
$$y = X\beta + e.$$

## Conclusion

The linear regression model gave the following results.

	B_hat	Standard Error	Lower 95%	Upper 95%
0	3.377802	0.061001	3.258057	3.497548
1	2.184620	0.060853	2.065164	2.304075

We can easily say that there is a direct proportion between happiness and GDP.



So, the final formulation is:

$$\text{Happiness Score} = 3.378 + 2.185 \times \text{GDP}$$