

**Koç University**  
**502 Advanced Data Analysis with Python**  
**Prof. David Carlson**  
**Homework 2**  
**Işık TOPÇU**

**Linear Regression Research Report**

*Do GDP and undernourishment have an effect on life expectancy?*

## 1. Data

For my research, I pulled World Bank API from 1900 to 2020 for Pakistan containing the variables below;

### Variables

- GDP Per capita (Independent variable)
- Prevalence of undernourishment (% of population) (Control variable)
- Life expectancy at birth, total (years) (Dependent variable)

Number of observations after NaN values: 19

## 2. Modelling

In my case, I generated a **multiple linear model** with the explanatory variable GDP and my outcome variable life expectancy. I added the causally prior control variable prevalence of undernourishment (% of population) as well. The model can run with an arbitrary number of variables although I just used two independent (one control) variables and a dependent variable. I wanted to see if GDP affects life expectancy because GDP per capita increases the life expectancy at birth through increasing economic growth and development in a country and thus leads to the prolongation of longevity. I also added the undernourishment control variable because nourishment is positively and biologically related to life expectancy and it is hypothesised that access to nourishment also affects GDP through the improvement of physical health and several researchers have found evidence that a healthier labour force can increase productivity and therefore economic growth.

## 3. Hypothesis

The null hypothesis states that there is no relationship between the two variables being studied (one variable does not affect the other). It states the results are due to chance and are not significant in terms of supporting the idea being investigated.

The null hypothesis stated that there should be no relation between GDP and life expectancy.

The alternative hypothesis stated that there is in fact a relation between GDP and life expectancy.

#### 4. Results

We can reject the null hypothesis if the t-values we found are bigger equal to the t-stat we found via the scipy module.

In my case, my t-values were;

The t-stat is approximately 2.10981.

- $t_{\text{GDP}} = 4.729635^1 \geq t$  Null hypothesis rejected.
- $t_{\text{undernourishment}} = 0.660613 < t$  Failed to reject the null hypothesis.

#### 5. Conclusions

Although we rejected the null hypothesis, I believe that a dataset with more observations and variables would've changed our results. Also, one must take into account that linear regression cannot usually reflect the real world. Furthermore, World Bank data is very limited. I will definitely try to extend my dataset via other sources to get a wider dataset to prove my hypothesis that there might be a correlation between GDP Per capita and life expectancy.

---

<sup>1</sup> Approximated. The actual t-value can be found on the model.