Question: How does the GDP per capita of a country affects the population distribution between rural and urban areas?
Hypothesis: Countries with higher GDP per capita have an economy more concentrated around service and industry sectors instead of agriculture. Service and industry sectors are located around urban areas. So higher GDP per capita implies more urban population percentwise.
Prediction: GDP per capita value and proportion of the urban population will have a simple linear relation.

To test the hypothesis, below data is gathered for all the countries around the world from the World Bank Database API:

1. GDP Per Capita (Current US$)
2. Urban Population (% of Total Population)

Since the World Bank Database does not have all the countries' data, list-wise deletion is performed to eliminate NaN values. A quick simple linear regression is applied to the data gathered to see any liner relation between the outcome variable (Urban Population) and the covariate (GDP Per Capita).

Since first results did not yield satisfactory regression estimates. In order to get more meaningful results, normalization and feature scaling is done to the variables. With simple normalization, satisfactory results are gathered. When regression model is run on the scaled data, below results are gathered. For further yields and more optimum results, advanced rescaling methods should be used. In this paper, advanced rescaling such as geometric or log-scaling is not applied, since it is far beyond the scope of this homework.

$$h(x_i) = \alpha + \beta x_i$$

Where $h(x_i)$ is the predicted response value. α and β values as well as standard error for beta are found as the following according to the model:

$$\alpha = 0.494$$
$$\beta = 0.773$$
$$SE(\beta) = 0.083$$

With given standard error, 95% credible interval for the β value is found as:

$$CI_{0.95} = [0.609, 0.936]$$

Result: In that case since 95% confidence interval does not contain 0, null hypothesis of not having any statistically significant relation between outcome variable (Urban Population) and the covariate (GDP per Capita) is rejected. There is a statistically significant linear relation between these two values according to the simple linear regression model.