

Multiple Linear Regression

(Djaouida Dahmani-Zaouche)

Exercise

In the following example, we will use multiple linear regression to predict the stock index price (i.e. the dependent variable) of a fictitious economy by using 2 independent/input variables:

- Interest Rate
- Unemployment Rate

Before applying linear regression models to predict the stock index price, you have to make sure that a **linear relationship** exists between the dependent variable and the independent variable/s (more on that under the *checking for linearity* section).

Here is the data-set that we'll use:

Year	Month	Interest_Rate	Unemployment_Rate	Stock_Index_Price
2017	12	2.75	5.3	1464
2017	11	2.5	5.3	1394
2017	10	2.5	5.3	1357
2017	9	2.5	5.3	1293
2017	8	2.5	5.4	1256
2017	7	2.5	5.6	1254
2017	6	2.5	5.5	1234

2017	5	2.25	5.5	1195
2017	4	2.25	5.5	1159
2017	3	2.25	5.6	1167
2017	2	2	5.7	1130
2017	1	2	5.9	1075
2016	12	2	6	1047
2016	11	1.75	5.9	965
2016	10	1.75	5.8	943
2016	9	1.75	6.1	958
2016	8	1.75	6.2	971
2016	7	1.75	6.1	949
2016	6	1.75	6.1	884
2016	5	1.75	6.1	866
2016	4	1.75	5.9	876
2016	3	1.75	6.2	822
2016	2	1.75	6.2	704

1. Build a `dataFrame` for our dataset, called `df`.
2. We want to apply linear regression model to predict the `Stock_Index_Price` with respect to the `Interest_Rate` and `Unemployment_Rate`. It is advisable to Check for Linearity before you execute a linear regression model; you may want to check that a linear relationship exists between the dependent variable and the independent variable/s. To perform a quick linearity check, you can use [scatter diagrams](#) (utilizing the `matplotlib` library. Write instructions to get the following charts:



3. Explain why a linearity relationship exists between the Stock_Index_Price and the Interest_Rate (resp.the Unemployment_Rate).

4. Perform a multiple linear regression for our dataset.

5. Consider again the previous question, normalize the Interest_Rate (resp.the Unemployment_Rate) by choosing a same range. Then perform a multiple linear regression.

6. Give the equation modelling the variation of the Stock_Index_Price with respect of the Interest_Rate and Unemployment_Rate by considering not normalized data, and the normalized one. What do you notice?
7. Why the normalization of data before applying multiple linear regression is recommended for big data?
8. Use python tkinter library to get the following window. To each button, associate an appropriate function.



