

Name: Siddharth Mohanty
Branch: EXTC
UID: 2019120038
Batch: B
Course: Data Analytics Lab
Experiment: 4

AIM: : Building Linear Regression model for given dataset.

Problem Statement:

Problem 1.1 - Creating Our First Model

We are interested in how changes in these variables affect future temperatures, as well as how well these variables explain temperature changes so far. To do this, first read the dataset `climate_change.csv` into Python.

Then, split the data into a training set, consisting of all the observations up to and including 2006, and a testing set consisting of the remaining years (hint: use `subset`). A training set refers to the data that will be used to build the model, and a testing set refers to the data we will use to test our predictive ability.

Next, build a linear regression model to predict the dependent variable `Temp`, using `MEI`, `CO2`, `CH4`, `N2O`, `CFC.11`, `CFC.12`, `TSI`, and `Aerosols` as independent variables (Year and Month should NOT be used in the model). Use the training set to build the model. Enter the model `R2` (the "Multiple R-squared" value):

Problem 1.2 - Creating Our First Model

Which variables are significant in the model? We will consider a variable significant only if the p-value is below 0.05. (Select all that apply.)

a) `MEI` b) `CO2` c) `CH4` d) `N2O` e) `CFC.11` f) `CFC.12` g) `TSI` h) `Aerosols`

Problem 2.1 - Understanding the Model

Current scientific opinion is that nitrous oxide and CFC-11 are greenhouse gases: gases that are able to trap heat from the sun and contribute to the heating of the Earth. However, the regression coefficients of both the `N2O` and `CFC-11` variables are negative, indicating that increasing atmospheric concentrations of either of these two compounds is associated with lower global temperatures.

Which of the following is the simplest correct explanation for this contradiction?

Exercise 3

I. Climate scientists are wrong that `N2O` and `CFC-11` are greenhouse gases - this regression analysis constitutes part of a disproof.

II. There is not enough data, so the regression coefficients being estimated are not accurate.

III. All of the gas concentration variables reflect human development - `N2O` and `CFC.11` are correlated with other variables in the data set.

Code & Output:

<https://github.com/51stDimension/EDAExperiments/blob/main/Experiment%204/EXP4.ipynb>

Conclusion:

All conclusions and final results have been added into the colab notebook mentioned in the above link.