Project Dataset 3: Pollution

S2 2024

Predicting the annual mean concentration of sulphur dioxide (per m^3)

This data set provides air pollution data for 41 US cities.

The annual mean concentration of sulphur dioxide, in micrograms per cubic metre, is a measure of the air pollution of the city. The question of interest here is what aspect of climate and human ecology as measured by the other six variables in the data set influences pollution in the city?

Your project is to investigate the association between the mean annual sulphur dioxide (SO_2) content of air in micrograms per cubic metre measured in a city, and other climate and human ecology variables that have been measured for the city.

After an exploratory analysis, you should derive:

A multiple regression model that provides the most accurate prediction of the mean SO_2 levels in the city. The analyses should consider the *effect* of the region, average annual temperature, number of manufacturing enterprises employing 20 or more people and other variables, and how outliers(if any), missing values(if any) and categorical variables should be treated, and the interpretation of the final model. **When deciding on a final predictive model, you must clearly show in your working how the final set of explanatory variables were selected (justify their inclusion in your model).**

A full list of variables, available for 41 cities, is below:

- *SO2*: average annual sulphur dioxide content of air in micrograms per cubic metre.
- *temp*: average annual temperature in Fahrenheit.
- *manu* : number of manufacturing enterprises employing 20 or more workers.
- *popul*: population size(1970 census); in thousands.
- *wind*: average annual wind speed in miles per hour.
- *precip*: average annual precipitation in inches.
- *predays*: average number of days with precipitation per year.
- *Region : the region in the USA*