## Global Warming



Global warming occurs when carbon dioxide (CO2) and other air pollutants and greenhouse gases collect in the atmosphere and absorb sunlight and solar radiation that have bounced off the earth’s surface. Normally, this radiation would escape into space—but these pollutants, which can last for years to centuries in the atmosphere, trap the heat and cause the planet to get hotter. Rising sea levels and an increased frequency of extreme weather events will affect billions of people.

In this problem, we will attempt to study the relationship between average global temperature and several other factors.

The file **GlobalWarming.csv** contains climate data from May 1983 to December 2008. The available variables include:

* *Year*: Calender year.
* *Month*: month of the Year.
* *Temp*: the difference in degrees Celsius between the average global temperature in that period and a reference value. **This is the Dependent variable.**
* *CO2*, *N2O*, *CH4*, *CFC.11*, *CFC.12*: atmospheric concentrations of carbon dioxide (CO2), nitrous oxide (N2O), methane  (CH4), trichlorofluoromethane (CCl3F; commonly referred to as CFC-11) and dichlorodifluoromethane (CCl2F2; commonly referred to as CFC-12), respectively.
  + CO2, N2O and CH4 are expressed in ppmv (parts per million by volume  -- i.e., 397 ppmv of CO2 means that CO2 constitutes 397 millionths of the total volume of the atmosphere)
  + CFC.11 and CFC.12 are expressed in ppbv (parts per billion by volume).
* *Aerosols*: the mean stratospheric aerosol optical depth at 550 nm. This variable is linked to volcanoes, as volcanic eruptions result in new particles being added to the atmosphere, which affect how much of the sun's energy is reflected back into space.
* *TSI*: the total solar irradiance (TSI) in W/m2 (the rate at which the sun's energy is deposited per unit area). Due to sunspots and other solar phenomena, the amount of energy that is given off by the sun varies substantially with time.
* *MEI*: multivariate El Nino Southern Oscillation index (MEI), a measure of the strength of the El Nino/La Nina-Southern Oscillation (a weather effect in the Pacific Ocean that affects global temperatures).

### **1 - Questions**

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 GlobalWarming.csv into R/Python.

1. Find the correlations between all the variables in the dataset except Date & Month. Which of the following independent variables is N2O highly correlated with (absolute correlation greater than 0.7)?

**CO2, CH4, CFC-12**

1. What is Multicollinearity? Does Multicollinearity exist in this Data?

Two independent variables are co related then its called Multicollinearity. Yes, it does

1. Which variables are highly correlated with the Dependent Variable? Is Co-relation god or bad for the Model? Why?

**N2O AND Co2 are highly correlated . Co relation is not good for model. It tends to make model imbalance towards the other side**

1. What will be the prediction of a baseline Model? What will be its SSE, more commonly known as SST/TSS?
2. Build a linear regression model to predict the dependent variable Temp, using CO2, N2O, CFC.12, and Aerosols as independent variables.Find the RMSE, SSE and MAPE for the above Model.What is the model R2 (the "Multiple R-squared" value & Adjusted R2) .
3. Calculate R2 value manually using the formulae for R2. Both should match.
4. Find the correlation between the actual & predicted values of Temp. Square this number and you should get the Multiple R2. This is another quick way to compute R2. ( R2= correlation(actual, predicted values)^2)
5. What is the Hypothesis test in case of a Linear Regression?

Hypothsis Test in Linear Regression and logistic

Here we assume that variable is not important

the variable is insignificant i.e., Beta (B) = 0

eg: y x1 x2

y = b0 +b1x1 + b2x2

x2 is given variable then we do b2 = 0 that means x2 is insignificant

Then we find p value then <5% then we reject null hypothesis and tell that the variable is important in the model

1. What do the P values represent? Do we look for low or High P values? Why?

The P value, or calculated probability, is the probability of finding the observed, p value is probaility or signficant value

plow H0 go Reject the null below 5%

phigh ho fly

1. What is an F Test in Linear Regression? Discuss.

The F-test in Linear regression is whether your linear regression model provides a better fit to the data than a model that contains no independent variables. R-squared tells you how well your model fits the data, and the F-test is related to it.