Predictive Analytics and Business Forecasting

**CO2 emission Analysis globally from 1920**

Here we are analyzing the dataset that contains the global Carbon dioxide emission data from 1920. As we see here the data has been collected monthly and all total there are 1200 observations.

The various columns in this dataset are CO2, Time, Date and CO2 Cumulative.

Initially I have uploaded the dataset from the local drive to the SAS using the below code.

proc import out= work.Proj

datafile= "/home/u49612406/sasuser.v94/CO2 Emission Data.csv"

dbms=csv replace; getnames=yes; datarow=2;

run;

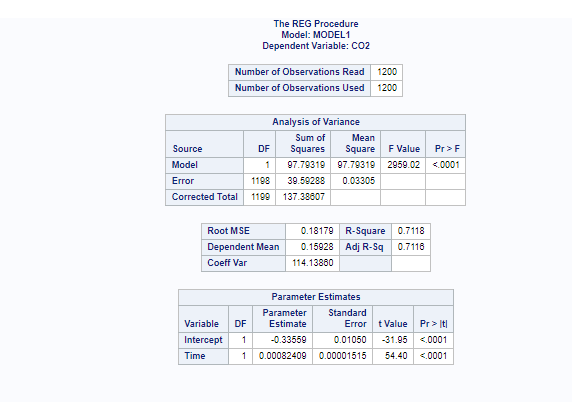
Then we have tried to run a simple linear regression model on the dataset where we have regression of CO2 on Time.

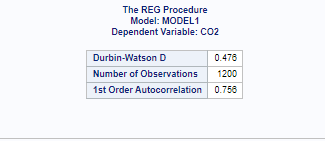
The below SAS code is used for this purpose.

proc reg data= work.Proj;

model CO2 = Time/clm cli dw;

run;





As we know This statistic indicates the percentage of the variance in the [dependent variable](https://statisticsbyjim.com/glossary/response-variables/) that the [independent variables](https://statisticsbyjim.com/glossary/predictor-variables/) explain collectively. R-squared measures the strength of the relationship between our model and the dependent variable on a convenient 0 – 100% scale

So, this model is capable of explaining 71.18 % of the total variance of the dependent variable.

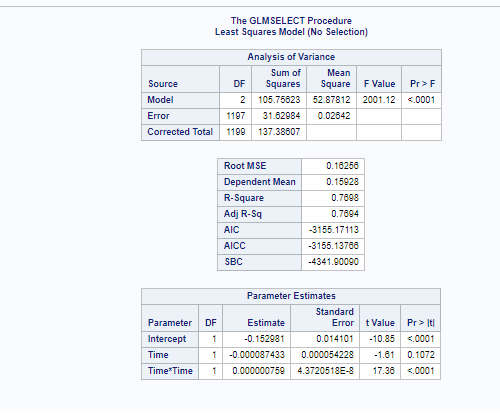
**CO2 = 0.000824\*Time - 0.33559**

2. Next we have tried to run quadratic regression on this dataset using the below SAS codes.

proc glmselect data=work.Proj;

model CO2=Time Time\*Time / showpvalues selection=none;

run;



Here we see that the R^2 has increased to 0.7698 and this quadratic regression model is capable of explaining 76.98 % of the total variance of the dependent variable.

CO2 = 7.59E-7\*(Time^2) - 8.7E-5\*Time -0.1529

3. Also

The Durbin Watson test conducted on the dataset gives a value of 0.476 which is from 0 to less than 2 and that indicates positive autocorrelation.

4. Hence the independency assumption is getting violated and so to improve the regression model, we have to use some kind of autoregressive process.

We will try to implement ARIMA model for this dataset now.