DEPARTMENT OF CHEMICAL ENGINERING
CH5440: MULTIVARIATE DATA ANALYIS
ASSIGNMENT 2

1. The following gases carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O) and Ozone (O3) in the atmosphere are implicated in increasing global temperatures, and are known as greenhouse gases. The concentration of these gases in the atmosphere and corresponding global average temperatures obtained from the EPA website (https://www.epa.gov/climate-indicators/weather-climate) between the years 1984 to 2014 is given in the Excel file *ghg-concentrations_1984-2014.xlsx* (units for different variables are also given in Excel sheet).

(a) Develop a multilinear regression model between global temperature (deviations) and concentrations of greenhouse gases using OLS. Is the global temperature positively correlated with increase in the concentration of these gases?

(b) Estimate the error variance in temperature measurements and confidence intervals (CIs) for all regression coefficients. Based on residual analysis, remove samples suspected of being outliers (one at a time) until there are no outliers.

(c) Improve the regression model obtained in step (b) by dropping unimportant (insignificant) variables (one at a time).

(d) The effect of different gases on the global temperature is expressed in terms of CO2 equivalents or global warming potential (GWP). Is it possible to make any inference regarding GWP of the gases from the regression coefficients? Compare the GWP obtained from regression coefficients to the values obtained over a 20 year time horizon: CO2 (1), CH4 (86), N2O (289).

Notes: Water vapour, which is present in significant amount is the atmosphere is also a greenhouse gas, but it remains almost constant and is relatively unaffected by human activity. CFCs/HCFCs which are also greenhouse gases are however being monitored only in recent years.

2. Consider the problem of developing a correlation between saturated pressure (P^{sat}) and saturated temperature T (boiling point). Data for saturated pressure for n-hexane in the temperature range 10 - 70 deg C are given in file vpdata.txt (first column is temp in K and second column is pressure in kPa).

(a) The Antoine equation is a modified Classius Clapeyron relation between ln P^{sat} and T and is given by

$$ln\ P^{sat} = A - \frac{B}{T + C}$$

Assuming that temperature measurements are noise-free and pressure measurements are noisy, use nonlinear regression to obtain estimates of parameters A and B and C by minimizing the squared difference between measured and predicted saturation pressures. Report the estimates obtained by using MATLAB function lsqnonlin and fmincon. (In order to provide good initial estimates of A, B and C use linear least squares estimates of A and B obtained by assuming C = 273 K)