**Exercise 8.2 - Model Diagnostics:**

You are given a file titled "TBCasesinUKUpdated.csv". Assume the simple linear regression model.

1. Obtain the estimate regression function.
2. Test at the 5 % level whether number of TB cases increases with year. Estimate the p–value of the test.
3. Examine and discuss if the 4 basic assumptions of linear regression hold for the model.

**Note:**

1. See Lab8.1FullSolution….docx document to get started. The following R code is what you will need for this example.

2. **But make changes to the names of variables etc.** Make sure your variable naming is meaningful.

install.packages("psych")

install.packages("nloptr")

library(psych)

library(car)

**#Obtain the estimate regression function**

HospitalCost1 <- read.csv("HospitalCost.csv", stringsAsFactors=FALSE)

# Build linear model

RegressedCost1<- lm(Cost ~ Size+Visibility+Admissions, data=HospitalCost1)

summary(RegressedCost1)

**#Interpret the results – p-value for predictor, p-value for F-test and Adj.R-squared**

# -------------------------------------------------------------------

plot(RegressedCost, which=1)

**#check the plot for constant variance, i.e. there is no specific pattern to the plot**

summary(lm(abs(res)~fitted(RegressedCost))) **#dirty test for a linear relationship**

# -------------------------------------------------------------------

res = resid(RegressedCost) **#get the residues**

summary(res) **#Look at how the residues are distributed and it gives an indication of normality**

plot(RegressedCost,which=2)

**#check normal plot, explain the nature of the plot – should be close to the normal line**

shapiro.test(res)

**#do a test for normality – high p-value means you cannot reject the null hypothesis that the residues are normally distributed**

# -------------------------------------------------------------------

durbinWatsonTest(RegressedCost)

**#do a Durbin Watson Test - high p-value indicates you cannot reject the null hypothesis that the residuals are not correlated**