MATH 4773 Laboratory 6: Multiple Regression part 4

In this laboratory you will extend the MLR of parts 1-3 to include residual analysis.

# Objectives

In this lab you will learn how to make use of R’s formula language to make models and use functions like predict() to examine model assumptions

### Tasks

Use RMD and knit into HTML. Place both completed documents on CANVAS

* Task 1
  + Download from CANVAS the zipped data files, “Dataxls”
  + Unzip the contents into a directory on your desktop (call it LAB)
  + Download the file “lab6.r”
  + Place this file with the others in LAB6.
  + Start Rstudio
  + Open “lab5.r” from within Rstudio (this is an exemplar file).
  + Using hash commenting and at the top of Lab 6 place the task number eg #Task 1
  + Go to the “session” menu within Rstudio and “set working directory” to where the source files are located.
  + Create the working directory by issuing the command getwd(): under #Task 1
* Task 2
  + Make sure you install the readxl package if you haven’t already.
  + On page 599 the text MS has an example worked in MINITAB using the data set KWHRS.xls
  + Answer Example 11.12 questions using R – see pg 598 by engaging with the data, R and the problem – or you will learn 0!!!
    - A)
    - B)
    - C)
    - D)
    - E)
    - F)
* Task 3
  + Residuals are used to check assumptions
  + Table 11.7 gives a good tabular summary of how to use residuals – see pg 606 How do you check to see if population variance is likely constant?
  + Read in the CLERICAL.xls data set and using the model y ~ x1+ … + x5 re-create the plots shown on page 608
  + What do they say? What do you conclude?
  + Recreate the table page 607 (Just the values under the column names) – place here:
  + Create a QQPLOT of the residuals. Place here!
  + Use library(s20x) and perform a normcheck() – place here (Make sure the graphic has shapiro -wilk in it)
  + Create the plot on page 610.
* Task 4
  + Create a function called myinfluential() that performs the analysis described on page 611 in the paragraph beginning “Several sophisticated .. “
  + This procedure is a “Jack-Knife” applied to regression. If is large then the datum removed is influential.
  + How could you decide that this measure is big or not?
  + Myinfluential() must output a list of all values of the appropriate measure with all those deemed significant.