MATH 4773 Laboratory 3: Multiple Regression part 1

In this laboratory you will learn the basics of practical and theoretical MLR or multiple linear regression. SLR (Simple Linear Regression) is the precursor to MLR and was covered in chapter 10 of MS. There will be a number of labs on this subject since the subject is vast in its applications and variations.

# Objectives

In this lab you will learn how to:

* Identify a linear model
* Identify MLR
* Apply matrix methods to the SLR
* Use R to carry out a SLR using matrix methods
* Use R for the case of two continuous predictors

### Tasks

This is to be done in R markdown. Knit into an html and place both rmd and html files 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 “lab3.r”
  + Place this file with the others in LAB3.
  + Start Rstudio
  + Open “lab3.r” from within Rstudio (this is an exemplar file).
  + Using hash commenting and at the top of Lab 2 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.
  + Copy and paste the working directory by issuing the command getwd(): under #Task 1
* Task 2
  + Make sure you install the readxl package.
  + Locate the file INSULATION.xls
  + Use read\_excel() to read the data into R, this function will already be available within the script lab3.r which you have opened in Rstudio.
  + Generate the working directory with getwd()
* Task 3
  + Using page 558 in MS write down the general form of a ML model in RMD:
  + The model is also called a general linear model – because ? (Hint: see footnote at the bottom of page 558)
  + Give an example of a non linear model from the same footnote and write its equation here:
  + Write down the model formula appropriate to the INSULATION data.
* Task 4
  + Write down the six steps used in analysing a Multiple Regression Model (see pg 559).
  + Write down the 4 assumptions of a Multiple Regression Analysis (see pg 560)
  + For the INSULATION data – how man independent variables are there?
  + Using the s20x library and the function normcheck() – what do you conclude? Paste all output from the function here:
* Task 5 ( This is about )
  + The INSULATION data set is small. Using R make a Y vector and X design matrix as outlined on page 561. Paste the vector and matrix here:
  + Find the MLR beta estimates in R from the formula on page 562 (which you will need to derive eventually) NB : You must use this formula in R.

Generate them in RMD.

* + Now write a function in R that will take any design matrix and Y vector and create a beta hat vector. Call the function mybeta(). Use it on the INSULATION data set and see if you get the same results as above – place function output here:
  + You can do this in R another way – use lm() and create y.lm = lm(…) paste summary(y.lm) here: Do the estimates agree?
* Task 6
  + Use table 11.3 page 563 and create a MLR model using matrix algebra in R. Copy and paste X and Y here:
  + Using matrix algebra in R make beta estimates – paste here:
  + Use mybeta() function and paste output here:
  + Use lm() and create a linear model object called y.lm2
  + Summarise y.lm2 – paste output here:
  + Do you get the same beta estimates?
* Task 7
  + Make a function that creates Y and X from data and then makes beta estimates and releases them in a list with the data to the command line and plots all the data (see pairs()).
  + You can be as creative as you wish – I will award more points for sophistication in the function. You will need to think about the dependent and independent variables what is on the dependent axis – if there are two x variables what plot would be appropriate – you can use contour() and other geoms in ggplot2 etc – impress me! Look at google!