MATH 4753 Laboratory 2

Introduction to R and Chapter 2

In this lab you will learn the basics of R. This program is free and you are encouraged to obtain a copy for your Mac, PC or Linux machine. Install it and then download and install R studio (this is a nice front end to R and is also free).

# Objectives

In this lab you will learn how to:

1. Use the empirical rule
2. Use the Chebyshev rule
3. Transform data to z values
4. Find outliers using z values

### Tasks

There are two front-ends to R that we will look at. Tinn-R and Rstudio.

We will use Rstudio for most of the course

All output made please copy and paste into this word file. Save and place in the dropbox when completed.

* Task 1
  + Download from CANVAS the zipped data files, “Dataxls”
  + Unzip the contents into a directory on your desktop (call it LAB2)
  + Download the file “lab2.r”
  + Place this file with the others in LAB2.
  + Start Rstudio
  + Open “lab2.r” from within Rstudio.
  + 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():
* Task 2
  + Find the file “EPAGAS.xls” inside LAB2
  + Open it in Excel
  + Save As type CSV(comma delimited) “\*.csv”
  + Use read.table(), read.csv() or the Rstudio menu to read the data into R (or any other method available), this function will already be available within the script lab2.r which you have opened in Rstudio.
  + Copy and paste the first six lines of the data using “head()” (use “courier new” font):
* Task 3
  + Make the object mpg, the number of miles per gallon vector.
  + If , then and Transform the mpg variable to z and verify these results.
  + Using z, find the values of mpg that are possible outliers.
  + Using z, find the values of mpg that defined as outliers.
  + Using the lattice package construct a dotplot with colors, Red=outlier, Blue=possible outlier. (NB – read the instructions in the lab2.r file for installing the package)
* Task 4
  + Make a boxplot of the mpg variable
    - Make the box black
    - Put a notch where the median goes
    - Put a title on the graph.
    - Make the plot horizontal.
  + Using Chebyshev’s theorem predict the proportion of data within 2 standard deviations of the mean of the mpg data.
  + Use R to calculate the exact proportion within 2 standard deviation of the mean.
  + Does Chebyshev agree with the data?
  + Now use the empirical rule, what proportion (according to the rule) of the data should be within 2 standard deviations of the mean?
  + How well does it correspond?
  + Is the Empirical rule valid in this case? Why?