ISE 3293/5013 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). If you do not have your own computer then you can use the machines on this level of the PHYSC building to do homework, assignments and projects.

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 D2L 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()`:

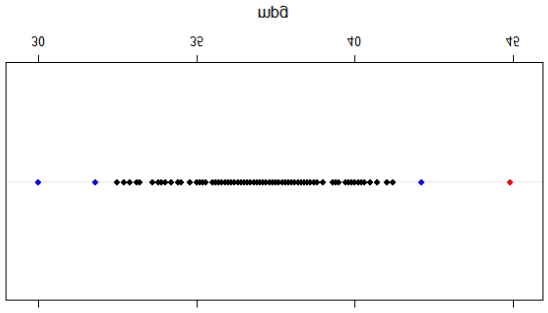
```
"F:/Google Drive - Saied/Courses/02 OU/11 Fundamentals of Enginee
ring Statistical Analysis/02 Labs/02 Lab 2
```

  -

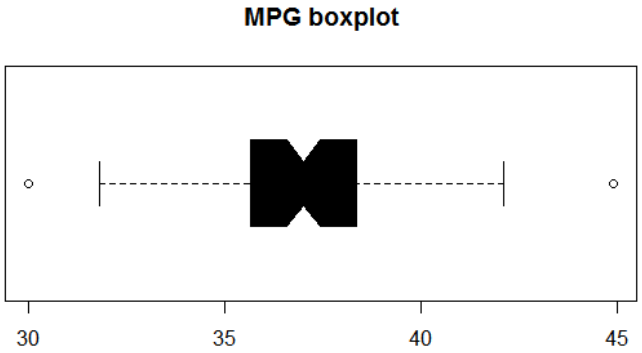
- Task 2
  - Find the file "EPAGAS.xls" inside LAB2
  - Open it in Excel
  - Save As type CSV(comma delimited) "*.csv"
  - Use `read.table()` to read the data into R (or any other method available), this function will already be available within the script lab2.r which you have opened in Rstudio.
  - Copy and paste the first six lines of the data using "`head()`" (use "`courier new`" font):

```
MPG
1 36.3
2 41.0
3 36.9
4 37.1
5 44.9
6 36.8
```

- Task 3
  - Make the object mpg, the number of miles per gallon vector.
  - If $z_i = \frac{(x_i - \bar{x})}{s_x}$, then $\bar{z} = 0$ and $s_z^2 = 1$. Transform the mpg variable to z and verify these results.
  - Using z, find the values of mpg that are possible outliers.
    ```
    30.0 42.1 31.8
    ```
  - Using z, find the values of mpg that defined as outliers.
    ```
    44.9
    ```
  - 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.

We Know that according to Chebyshev that if k(SD) = 2, then proportion of data within k deviations is greater than or equal to $\frac{1}{1-k\text{^}2}$. So, $1-\frac{1}{2^2}=1-\frac{1}{4}=\frac{3}{4}$. So, above 75% proportion of the mpg data falls inside 2 standard deviations of the mean of mpg data.

- Use R to calculate the exact proportion within 2 standard deviation of the mean.

```
0.96
```

- Does Chebyshev agree with the data?

```
Yes 96% is greater than 75% obtained my Chebyshev's theorem
```

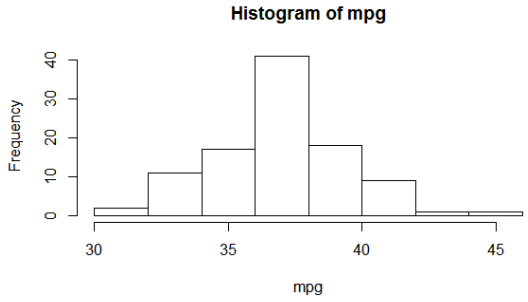
- Now use the empirical rule, what proportion of the data should be within 2 standard deviations of the mean?

```
According to Empirical rule approximately over 95% of the data lies within 2 standard deviation
```

- How well does it correspond?

```
Yes 96% is very close to 95% approximation of empirical rule.
```

- Is the Empirical rule valid in this case? Why?




```
By seeing the distribution from the histogram we can say that it is unimodal around the mean value. Then data is pretty symmetric on both sides of then mean. So, we can say that empirical rule is valid.
```

-