Data Science

Lab. Sheet 4 – Statistics & Data Import

Exercise 1
Set up a matrix to store the following data:
No_of_texts = [1, 2, 4, 1, 0, 6, 8, 9, 1, 1]
Time_on_phone = [5, 9, 12, 3, 10, 20, 24, 18, 4, 5].

You are now required to implement the code in Statistics 1, Statistics 2 and from data visualisation to

(i) calculate the following for each variable:
Measures of central tendency: Mean, Median, Mode
Measures of spread: Range, Interquartile Range, variance and standard deviation.

(ii) for the full data set: investigate relationship between the two variables, by
(a) creating a scatterplot;
(b) determining the covariance and the correlation.

(iii) what do the covariance and correlation values tell you about the relationship between the data. Research online the significance of correlation values and use this information to inform your conclusion.

Exercise 2

Read the first section of the "Getting Data" tutorial and use this to import the tab delimited "stocks" file on Moodle into suitable matrices.

You are now required to analyse the data with regard to

(i) Central tendency of data and spread over time
(ii) Suitability of a linear regression model to model the data and to predict the future stock price. Use the model to predict the stock price one month into the future. (You will need to investigate how to fit a linear model and to assess its suitability for prediction).
(iii) for the full data set: investigate relationship between pairs of variables by creating scatterplot and determining the covariance and the correlation. Comment on the significant of the correlation values and the information this provides as to future performance.