Project

# Question 1 of 2

1. For this question, the data are contained in the file Prices.txt, available on Blackboard. The data consist of the daily share prices of one of the popular companies in the UK, in dollars($). The data was obtained by a data scientist for purely research purposes. The sample size is 520.
   1. Read the data into a data frame called Prices in R
   2. Assuming the data scientist told you that the first 300 rows are based on 2019 share prices, and the next 220 rows are prices from 2020. Incorporate this piece of information into a second column in your data frame, called Price$year. For simplicity, you may wish to use rep command.
   3. Draw two box plots; one for the price distribution in 2019 and another for the price distribution in 2020. Comment on the main features of the plots. Do you think that the distributions are symmetric? I strongly suggest that you put the two box plots on the same axes to facilitate comparison.
   4. Put the 2019 price values into a variable called Prices2019.
   5. Suppose *x*1*, · · · , x*300 correspond to the 2019 data. Suggest a suitable transformation say, T such that the transformed data will be normally distributed *N* (*µH, σ* ). That is, *hi* = *T* (*xi*). Fit a normal distribution to the transformed 2019 data so as to obtain the estimates of the parameters of *hi*.
   6. Plot a density histogram of the transformed price values from 2019, and superimpose the p.d.f. of the fitted normal density on the same plot. Comment on the goodness of fit.
   7. Use the fitted model to estimate the probability that a randomly selected share price in 2019 is between 40 and 70. You may which to compare the estimated probability with the corresponding empirical probability(this last sentence is optional).

# Question 2 of 2

1. The data for this question are contained in the file distance.txt, available on Blackboard. The data consist of the distance, in miles, 100 randomly selected UCL staff commute to work each day.
   1. Read the data into R
   2. Draw a histogram of the data.
   3. Fit any distribution to the data that provides an adequate fit(You may consider fitting more than one distribution for comparison purposes. However, this will not give you extra mark). You may use the R package fitdistrplus
   4. Superimpose the fitted PDF on top of the histogram. Briefly describe how the fitted model match with the histogram.
   5. According to the Kolmogorov Smirnov test, do you think that the fitted distribution provided an adequate fit. Note that Kolmogorov Smirnov test can be done in R by using the command ks.test
   6. Construct a 95% confidence interval for the mean distance in miles UCL staff travel each day to work.