Tutorial 4

1. In many manufacturing processes the term work-in-process (often abbreviated WIP) is used. In a book manufacturing plant the WIP represents the time it takes for sheets from a press to be folded, gathered, sewn, tipped on end sheets, and bound. The data set `wip.txt` represents samples of 20 books at each of two production plants and the processing time (defined as the time in days from when the books came off the press to when they were packed in cartoons) for these jobs. There are two variables in the this data set. They are `time` and `plant`.

   (a) For each of the two plants, using R to compute the following descriptive statistics: mean, median, first quartile, third quartile, minimum, maximum, the range, interquartile range, variance and standard deviation.

   (b) Draw a histogram and box plot for the processing time for each of the two plants using R

   (c) Based on (a) and (b), are there any differences between the processing times of the two plants? Explain.

   (d) Repeat (a) and (b) using Python.

2. The director of a training program for a large insurance company wanted to know if the two tests, Test A and Test B, are correlated. Thirty trainees sat for both tests. The test scores are recorded in the file `testscores.txt`. There are three variables in the this data set. They are `A`, `B` and `gender`.

   (a) Using R, draw scatter plots for the two test scores for all the trainees, with a different symbol for different gender.

   (b) Using R, draw separate scatter plots for the two test scores, one for the male group and one for the female group.

   (c) Based on the plots in (a) and (b), is there any relation between the two test scores? Explain.

   (d) Repeat (a) and (b) using Python.

3. One of the major measures of the quality of service provided by any organization is the speed with which the organization responds to customer complaints. During a recent year a company got 50 complaints. The file `furniture.txt` which consists of the number of days between the receipt of the complaint and the resolution of the complaint.

   (a) Use R to find the 20% trimmed mean and 20% Winsorized mean. What can you say about the "central" location of the distribution?

   (b) Use R to find the interquartile range and the median absolute difference (MAD). What can you say about the "variability" of the distribution?

   (c) Repeat part (a) and (b) using Python.