First Group Assignment

This is our first of two homework assignments (in addition to two reading assignments). The value of this assignment is 5% of the course. Each group will also receive a different customized dataset to be analyzed using the R language and environment for statistical computing and graphics. The dataset is based on the electricity consumption data studied in the course project.

Please complete the tasks described below and submit an electronic copy of your solution to Amir (at sayaghou@sfu.ca) by October 6, 2018.

The dataset determines a *multivariate time series* by describing various features over time, including the following ones:

A. Global_active_power

B. Global_reactive_power

C. Voltage

D. Global_intensity

For the dataset assigned to your group, complete the following tasks:

1. Compute the *arithmetic* and the *geometric mean*, the *median*, the *mode* and the *standard deviation* for features A and B respectively.

2. Compute the *correlation* between each of the four features A, B, C and D using Pearson's correlation coefficient as defined below.

   If we have a series of *n* measurements of two discrete random variables *X* and *Y*, written as x_i and y_i for *i* = 1, 2, ..., *n*, then the *sample correlation coefficient* can be used to estimate the population Pearson correlation *r* between *X* and *Y*. The sample correlation coefficient is a measure of the linear correlation between *X* and *Y*, and can be written as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

   where x̄ and ȳ are the *sample means* of *X* and *Y.*

   The following command in R allows to calculate Pearson's correlation.

```
cor(var1, var2, use = "", method = "")
```

3. For features A and B compute the *min* and *max* values on weekdays and weekend days respectively.

   The command in R to read a ".txt" file is the following one:

   ```
   read.table(fileName, header = , sep = "")
   ```

   In order to extract specific days from a time series you will need this command:

   ```
   as.POSIXlt(date, format = "")
   ```

   (See also: https://stat.ethz.ch/R-manual/R-devel/library/base/html/strptime.html)