**BA4318 Fall 2018 Sample Midterm Exam**

Four athletes are training for the Olympic 100 meter dash for the past 2 years. Every day after training, they do a test run. You keep records in terms of milliseconds.

You consider less than 10.000 milliseconds as athlete-level performance. All team is less than 10.000 milliseconds lately. However, you need to select two of for the competition team.

* The obvious candidates are Joe and Bill who have been improving steadily.
* On the other hand, John has been consistently beating them for the past few weeks.

You just cannot select three. You are also suspecting John started using performance improvement drugs. However, you cannot force him to take advanced drug tests unless you can validate your suspicion.

You can use statistics to check. **If John’s late improvement cannot be predicted easily by his past performances, then it should be attributed to something else.**

Your data set is in the file Runners.txt, which is a tab-separated data set. The first column is the run number and the second column is the date. The remaining columns are the run durations in milliseconds. All durations are integers. There are 739 data points for each athlete. There is no missing data.

**Questions**

1. Use statistical estimation techniques to re-estimate the last 22 runs, based on the previous 717 runs and calculate the error terms. (70 points)
   1. Use at least two different methods to estimate. One of them should take into account seasonality and one should not. When there is a parameter in the method you may need to run the program multiple times, changing the parameter (i.e. alpha) to see the better results.
   2. Calculate the sum of errors for all four athletes for the two (or more) methods and select the method with the smaller error.
2. Based on the selected “better” method, compare the error terms for all four athletes. (30 points) If John’s error term is much larger (%50 or more) than the average error of all other three then you can validate your suspicion.

Note: You should define some functions in your code.

* Use variables to record the error terms for each output, and variables for the sums of the errors, averages of the errors, etc. You may use many variables or tuples. All are OK.
* Define separate functions for the estimation methods, the parameters being the data frames as an input, and the error as an output (i.e. return the error).
* Define a function to compare the errors for the athletes.
* Do not forget to have a single line output stating the result.