

**Homework # 10**

*Due on 05/06/2020 at 1:00 pm*

1. The data on national track record for women for Helsinki 2005 are in the file `TrackWomen.dat`. Note that the file is *tab-delimited* and has a header with the variable names. Note also that the times for the events 100 m to 400 m are measured in seconds, while 800 m to Marathon events times are measured in minutes. Convert the data to speeds as we did in previous homework assignments.
  - a) Carry out a hierarchical cluster analysis using *average* linkage method. Plot the dendrogram. Looking at the chart and using largest change in distance rule of thumb, how many clusters should you use?
  - b) Carry out a cluster analysis using *centroid* linkage method. Plot the dendrogram. Find the distance between the centroids of the two-cluster solution.  
Note: For this homework you might want to take a look at the function `cutree` which will allow you to classify each observation to the cluster it belongs to.
  - c) Using two clusters from part a), label the observations from one cluster as Group 1 and the observations from the other cluster as Group 2. Calculate the linear discriminant function between the groups. Does it have a good interpretation?
  - d) Carry out a *k*-means cluster analysis with 2 clusters. Is the result the same as in part c)?
  - e) Finally, repeat part d) but *on the original* data matrix before converting to speeds. Are the results identical to part d)?