HUDM 6122
Spring 2020

Homework # 7

Due on 04/01/2020 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 has a header with the variable names, and 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.

   a) Obtain the sample correlation matrix $\boldsymbol{R}$ and report its eigenvalues and eigenvectors.

   b) Report the first two principal components for the standardized variables, and the cumulative percentage of the total sample variance explained by the first two components.

   c) Interpret the two principal components, that is, which variables dominate them?

   d) Rank the nations based on their score on the first principal component. Does this ranking correspond to your intuitive notion of athletic excellence for the various countries?

   e) Convert the data to speeds, measured in meters per second. Note that the marathon is 26.2 miles, or 42,195 m long.

   f) Perform a PCA on the covariance matrix $\boldsymbol{S}$ of the speed data from part e).

   g) Compare the results with parts b) – d) above.

   h) Which analysis do you prefer? Why?