Exercise 7.2

Principal Component Analysis on the 'Fitness' dataset

the 'Fitness' dataset is presented in the **SAS Manuals on the Reg procedure**. It contains Data on Physical Fitness, and the "measurements were made on men involved in a physical fitness course at N.C. State Univ. The variables are Age (years), Weight (kg), Oxygen intake rate (ml per kg body weight per minute), time to run 1.5 miles (minutes), heart rate while resting, heart rate while running (same time Oxygen rate measured), and maximum heart rate recorded while running. ***Certain values of MaxPulse were changed for this analysis".

The data are available in the file fitness.txt. Read in the data in R.

A. Analysis on all variables

1. In a principal component analysis, should we analyze the correlation- or the variance-covariance matrix? Why?
2. Look at the eigenvalues and related plots (i.e. scree- and variance explained plots). How many components should we use?

B. Exlude Age and Weight

3. And then repeat the above analyses.

C. Condition on Age and Weight

4. Compare the partial correlations with correlations in **B**. Compare the eigenvalues and vectors with those from the results in B!
5. Are there big differences? If not: what can we then say about Age and Weight?

D. Partial correlations based on Weight (by hand!)

6. Reproduce some of the partial correlations, the test statistic for assessing whether the true partial correlations are zero and find the associated p-values!