Homework 3 Due date: March 23

1. Exercise 3.18, 8.4, 8.6 in the book of Johnson and Wichern.
2. Given the car price data (https://www.kaggle.com/hellbuoy/car-price-prediction),
   (a) Determine the sample principal components and their variances with all the continuous variables except for *price* (*wheelbase, carlength, carwidth, carheight, curbweight, enginesize, boreratio, stroke, compressionratio, horsepower, peakrpm, citympg, highwaympg*).
   (b) Please choose the proper number of principal components that can best explain the variation of the 13 variables in (a) according to the variance explained and the scree plot. Interpret the principal components.
   (c) Calculate statistical distances $\left(\boldsymbol{x_j} - \bar{\boldsymbol{x}}\right)' S^{-1}\left(\boldsymbol{x_j} - \bar{\boldsymbol{x}}\right)$ for all observations with the 13 variables in (a). Determine the proportion of observations $\boldsymbol{x_j}$ falling within the distance 1.4^2.
   (d) For the vectors $\boldsymbol{v_j}$ formed by the first two principal component scores, Sketch the contour $\left(\boldsymbol{v_j} - \bar{\boldsymbol{v}}\right)' \Lambda^{-1}\left(\boldsymbol{v_j} - \bar{\boldsymbol{v}}\right) = 1.4^2$, where Λ is a diagonal matrix with the first two eigenvalues of S in the diagonal. Highlight the points selected in (c) with the red color. Discuss the consistency/inconsistency of the contour and the highlighted points.
   (e) Compute the correlation coefficient between the first two principal component scores and the variable *price*.