STAT 636, Fall 2015 - Assignment 6
Due Wednesday, Nov. 11, 11:55pm central
Online Students: Submit your assignment through WebAssign.
On-Campus Students: Email your assignment to the TA.

1. For the stock price data in textbook Table 8.4 ($n = 103$, $p = 5$):

   (a) Using the sample covariance matrix $\mathbf{S}$, find the sample principal components $\mathbf{y}_1 = \mathbf{X}\mathbf{a}_1$, $\mathbf{y}_2 = \mathbf{X}\mathbf{a}_2$, ..., $\mathbf{y}_p = \mathbf{X}\mathbf{a}_p$, where $\mathbf{X}$ is the $n \times p$ matrix of stock prices. Print the first five rows of $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p]$.

   (b) Determine the proportion of the total sample variance explained by the first three principal components. Interpret these components.

2. For the census-tract data in Table 8.5 ($n = 61$, $p = 5$), convert median home value to be measured in thousands rather than hundred thousands (multiply that column by 100).

   (a) Carry out two principal component analyses, one based on the sample covariance matrix and the other based on the sample correlation matrix, using the modified data. For each analysis, report the proportions of total sample variance explained by the first three PCs as well as the correlation between these PCs and the individual variables.

   (b) Interpret the PCs. Comment on how the interpretations differ between the two analyses. Which of the two would you recommend using, and why?

3. For the bull data in Table 1.10 ($n = 76$, $p = 8$):

   (a) Perform PCA on the standardized variables (i.e., based on the sample correlation matrix), excluding the variable `Breed`. What proportions of variability do the PCs explain?

   (b) Report a scree plot and comment on how many PCs are minimally necessary to adequately represent $\mathbf{X}$.

   (c) Interpret the first two PCs.

   (d) Report a scatterplot of the first two PCs, with the points color-coded by `Breed`. Can you distinguish groups representing the three breeds of cattle? Are there any outliers? If so, inspect their original variables and comment on what makes these bulls unusual relative to the other bulls.