Statistics 407 Homework 1

Due date: Wednesday, September 5, 2012, in class.
Although you may work together, you should hand in an individual solution.

1. Do the calculations for this question by hand, using the following data matrix:

$$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 2 & 3 \\ 3 & 4 \\ 1 & 3 \end{bmatrix}$$

   (a) What are the values of p, n, x_{32} and $\mathbf{x}_3$?

   (b) Find the mean vector, $\bar{\mathbf{X}}$, the sample variance-covariance matrix, $\mathbf{S}$, and the sample correlation matrix, $\mathbf{R}$.

   (c) What is $\mathbf{X}'$. (This is the transpose of $\mathbf{X}$.)

   (d) Compute $\mathbf{Z}$, the matrix of standardized scores for these data.

   (e) What are the values of z_{32} and $\mathbf{Z}_3$?

   (f) Construct a scatterplot of $\mathbf{x}_1$ against $\mathbf{x}_2$. Construct a scatterplot of $\mathbf{z}_1$ against $\mathbf{z}_2$. Explain the difference between these two plots.

2. Download the Ames housing prices data from the class web page, and answer these questions using a computer.

   (a) What are the values of p, n?

   (b) Report the summary statistics for this data: the variable means, standard deviations, and the sample correlation matrix. (Remember: Just use the numeric variables for this.)

   (c) Which variable has the highest mean? Which variable has the smallest variance? Which two variables are most strongly correlated?

   (d) Construct a scatterplot matrix of the data. List three things that you learn about the pairwise relationships in this data from the plot, that is different from what you would infer from the correlation matrix alone.

   (e) Now compute the SalePrice means by neighborhood. Which neighborhood has the highest average SalePrice?