**STAT505 Assessment #2**

1. Water and mosquito fish samples were collected at a sample of 96 sites in the Florida Everglades in the spring of 1996. The resulting data are stored in the file "marsh.txt", available on ANGEL. The data are provided in the following columns: 1) station code (alpha-numeric), 2) longitude, 3) latitude, 4) total mercury, 5) methyl mercury, 6) turbidity, 7) total phosphorus, 8) mercury in fish.

   These data may be input into SAS using the following code:

```
data marsh;
infile "v:\marsh.txt";
input station $ lon lat thg mehg turbid tpw fish;
run;
```

   The "$" sign after station is used to indicate that this is an alpha-numeric variable.

   (a) Produce histograms for total mercury, methyl mercury, turbidity, phosphorus, and fish.

      i. Should a normalizing transformation be applied to any of the variables? If so, what transformation (if any) is most appropriate (e.g., square root, quarter root, log) for each variable.

      ii. If transformations are appropriate, produce histograms for each of the transformed variables.

   (b) Using the transformations from part (a), produce a matrix of scatter plots for total mercury, methyl mercury, turbidity, phosphorus, and fish. Are there any outliers? In terms of the strengths of linear relationship, how would you interpret these plots?

2. (*adapted from J&W Exercise 4.24a*) (the data for this problem is in the file "companies.dat") Exercise 1.4 contains data on three variables for the world's 10 largest companies as of April 2005. For the sales ($x_1$) and profits ($x_2$) data, construct a $Q-Q$ plot. Do these data appear to be normally distributed? Explain.

3. Obtain the 95% prediction ellipse for each of the following covariance matrices:

$$\Sigma_1 = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{bmatrix},$$

$$\Sigma_3 = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 2.0 \end{bmatrix}, \quad \Sigma_4 = \begin{bmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{bmatrix}.$$

   You may assume $\mu_1 = \mu_2 = 0$ in each case. With $\Sigma_1$ as a reference, comment on the comparison with

   (a) $\Sigma_2$?

   (b) $\Sigma_3$?

   (c) $\Sigma_4$?