STAT505 Assessment #3

1. (*data from J&W Exercise 1.6*) The data in "air.dat" are 42 measurements on air-pollution variables in Los Angeles. The columns are x_1 = wind, x_2 = solar radiation, x_3 = CO, x_4 = NO, $x_5 = NO_2$, $x_6 = O_3$, and x_7 = HC.

   (a) Using notation consistent with our lessons, explain briefly what X_{ij} and μ_j represent in terms of the variables here. Similarly, explain briefly what σ_{jk} represents. What does it mean if $\sigma_{jk} = 0$?

   (b) Considering only solar radiation, CO (carbon monoxide), and NO_2 (nitrogen dioxide), can these variables reasonably be assumed to be multivariate normal? Justify your answer with appropriate plots.

   (c) Provide a matrix of scatterplots for the three variables above. Also, report the numeric correlation for each pair of variables. Is any pair significantly correlated? Answer this with separate confidence intervals of correlation. Use a Bonferroni adjustment for multiplicity so that your confidence for all intervals simultaneously is 95%.

2. Measurements of biochemical oxygen demand (Y_1) and suspended solids (Y_2) were obtained from the discharge of $n = 11$ municipal wastewater treatment plants into the rivers of Wisconsin. The values are recorded in the data set "water.dat" and can be read into SAS with the following code. The "firstobs=2" tells SAS to skip the first row, which includes labels.

```
data water;
infile 'v:\water.dat' firstobs=2;
input trt y1 y2;
run;
```

   Assume that these data are sampled from a bivariate normal population with mean vector μ and covariance matrix $\mathbf{\Sigma}$.

   (a) Find the sample mean vector and the sample covariance matrix.

   (b) Find 95% confidence intervals for the population means using

      i. One at a time multiplier

      ii. Bonferroni multiplier

      iii. Simultaneous confidence region multiplier

      Provide an interpretation of each of these intervals.

   (c) Compute the sample correlation r between oxygen demand and suspended solids.

   (d) Test $H_0 : \rho = 0$ against $H_a : \rho \neq 0$ at the $\alpha = 0.01$ level, where ρ is the population correlation between oxygen demand and suspended solids. What are your conclusions?

   (e) Give a 95% confidence interval for ρ.