CE 369 Applied Geostatistics
Homework Assignment #1 Normality and Correlation

For this homework assignment, we testing the normality and correlations in a 'real' data set. The data you will be using are hydrochemical samples taken from groundwater monitoring wells from a landfill leachate contaminated aquifer in Clinton County, New York. The data consists of position (Easting, Northing), 11 constituents (e.g. pH, SO_4, etc.) and type. The Type parameter is a categorical data defined as 1, 2 or 3 where 1 represents clean monitoring locations, 2 as along the fringe, and 3 as contaminated monitoring locations. Using the JMP or SAS software, please answer the following questions:

NOTE: You can copy and paste directly from JMP/SAS into word and annotate by hand/text so that you don't need to re-type all the information.

1) Test the normality of each sampled parameter. Which parameters appear to come from a normal distribution and which do not? Please give the statistics, bar plots, quantile plots, and normal quantile plots associated with each parameter.

2) Transpose the data using a natural log function and test the normality of the transposed data. Which parameters come from a normal distribution after transformation? Please give the statistics, bar plots, quantile plots, and normal quantile plots associated with each parameter. How do you deal with negative data during a natural log transformation (e.g. Eh)?

3) What should we do with the remaining sample parameters? Can we analyze them?

4) Find regression models for the following pairs of data: Fe and pH, Mn and pH, Mn and Fe, TOC and BOD_5, and Ln(COD) and Ln(BOD_5). What are the regression models? What is the difference between R^2 and R^2 Adj? Which of these regressions are significant?

5) Please comment on the quality of the TOC by BOD_5 regression.