Long Problem Set 4

For each problem below, complete any requested calculations and answer any accompanying questions. Your responses are evaluated on the appropriateness of your approach and the insightfulness of your analysis. Be sure to consider significant figures when interpreting the output from R (which, as with a calculator, often ignores such niceties).

You will share your answers to these problems in two ways: a text document that contains your written responses to the questions, and a `.R` script file that contains your code (use comments to separate your code by problem). Save your script file using `yourlastname_LPS02.R` as a file name and share it with by email. You may use any program you wish for your text document.

Your answers to the following questions are due by 4:00 pm on Tuesday, September 25th.

1. The data at this link (http://www.rsc.org/images/CO2_methods_tcm18-57755.txt) reports results for the determination of CO_2 by six different methods. The data itself uses Na_2CO_3 as a reference standard; presumably, a portion of the standard was treated to release the CO_2, which subsequently was determined and reported as % w/w CO_2 in the sample. Complete an analysis of variance on this data and identify sources of difference between the methods at $\alpha = 0.5$. The file carbonDioxide.RData contains vectors with the results for each method and a data frame that gives the concentration of CO_2 in the first column and the analytical method in the second column.

2. One important use of an analysis of variance is the ability to use the results to classify samples (a topic to which we will return later in the semester). For example, the file Pottery.RData has five objects that give the concentrations of Al, Ca, Fe, Mg, and Na (as %w/w metal oxide) in pottery shards collected at four different sites, which are identified as the factors "A", "C", "I", and "L" in the object Site. If there are significant differences in the concentration of a metal in pottery shards collected at different sites, then these differences might serve as characteristic markers that can help classify pottery shards of unknown origin. Use a separate one-way analysis of variance for each metal, using $\alpha = 0.05$, and determine if there is a way you can determine the origin of a pottery shard based on the concentration of one or more of the metals. Your answer should explain clearly how you could identify whether a pottery shard is from site "A", "C", "I", or "L".