Homework 9 (20.5 pts.)
due March 28

A reminder – Please do not hand in any unlabeled or unedited SAS output. Include in your write-up only those results that are necessary to present a complete solution (what you want the grader to grade). In particular, questions must be answered in order (including graphs), and all graphs must be fully labeled (main title should include the question number, and all axes should be labeled). Don't forget to put all necessary information (see course policies) on the first page. Include the SAS input for all questions at the very end of your homework; this could be important even though it won't be graded. You will often be asked to continue problems on successive homework assignments so save all your SAS code.

1. (7 pts.) For this problem, the idea is to demonstrate the similarity between regression with dummy variables and ANOVA and to gain practice in converting between the various parameters associated with ANOVA. To do this run the SAS code stat512hw9.sas.
   (a) Compare the ANOVA table and parameter results from the GLM analysis and Parameterization #1. What do the coefficients associated with X_1 and X_2 (i.e., b_1 and b_2) estimate in terms of treatment means? What constraint system does this parameterization correspond to? Why? If the constraint system does not match the parameter results from SAS, calculate the appropriate parameters to confirm your choice.
   (b) Compare the ANOVA table and parameter results from the GLM analysis of Parameterization #2. What do the coefficients associated with X_1 and X_2 (i.e., b_1 and b_2) estimate in terms of treatment means? What constraint system does this parameterization correspond to? Why? If the constraint system does not match the parameter results from SAS, calculate the appropriate parameters to confirm your choice.
   (c) Calculate $b_0 + b_1$ for both parameterizations and show that the answers are the same. What does this quantity estimate in terms of the treatment means (i.e., why are they the same)?
   (d) Calculate $b_1 - b_2$ for both parameterizations and show that the answers are the same. What does this quantity estimate in terms of the treatment means (i.e., why are they the same)?
   (e) Calculate the parameters from the cell means model from (using the definition of τ_i):
      i. the parameters in part a)
      ii. the parameters in part b)
   (f) From the parameters in the cell means model (e), calculate the parameters from the factor effects model with the following constraint systems:
      i. conceptual constraints
      ii. SAS constraints
2. (13.5 pts.) The next question uses the dataset from Problem 16.11 (Filling machines) in the book (CH16PR11.DAT).
   (a) Using the assumption of constant variance, what are the confidence intervals for the mean amount of detergent that is filled from (no multiplicity correction should be used): Be sure to write down your answer separately or clearly indicate your answer being sure to indicate which confidence interval is for machine 3 and which is for machine 4.
      i. machine 3
      ii. machine 4

(b) Use the Tukey multiple comparison method to determine which pairs of agents differ significantly. Summarize the results using a graphical method (lines). Be sure to summarize your results, that is state which machines are the same and which are different. If you just had the output from cldiff (confidence intervals), but no '***', how would you determine if the two means were the same or different?

(c) Is there a difference between the results of a) and b). Please explain your answer.

(d) Assuming that the electricity required to run the machines is the least in machines 4 and 2. Which machine would the company prefer to be used the most? Why?

(e) Machines 1 and 2 were purchased new five years ago, machines 3 and 4 were purchased in a reconditioned state five years ago, and machines 5 and 6 were purchased new last year. For the machines that were purchased five years ago, suppose you want to compare the average of the machines purchased new with the average of the machines purchased reconditioned. Use the `estimate` and `contrast` statements in `proc glm` to test the appropriate hypothesis. Report the estimated value of this contrast with its standard error; state the null and alternative hypotheses, the test statistic with degrees of freedom, the p-value, the decision and your conclusion.

(f) Check the assumptions using the appropriate plots. Explain in detail whether or not each assumption appears to be substantially violated.

(g) Check the homogeneity assumption quantitatively using the Levine test. Be sure to state the null and alternative hypotheses, the test statistic with degrees of freedom, the p-value, the decision and your conclusion.