Homework 2 (16 pts. + 0.5 BONUS) due Jan. 24

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. (11 pts.) For this problem, use the "plastic hardness" data described in the text with problem 1.22. (CH01PR22.DAT) Make sure you understand which column is X and which is Y and read in the data accordingly.
   (a) Plot the data using PROC GPLOT. Include a regression line on the plot (i = rl). Is the relationship approximately linear?
   (b) Give
      i. the linear model used in this problem
      ii. the estimated regression equation.
   (c) Run the linear regression to predict hardness from time. Give the complete ANOVA table for the regression.
   (d) Describe the results of the significance test for the slope. Give the hypotheses being tested (H_0, H_A), the test statistic with degrees of freedom, the p-value, the decision and your conclusion in sentence form.
   (e) Explain why or why not inference on the intercept is reasonable (i.e. of interest) in this problem.
   (f) Plot the 95% bounds (confidence band) for the mean (9.2: use i=rlclm on the symbol1 statement and the 95% bounds for the individual observations (using i=rlcli). Note: If you are using 9.2, you also need to calculate the % to use. Which one of these is wider? Why?
   (g) Give an estimate of the mean hardness that you would expect after 36 and 43 hours; and a 95% confidence interval for each estimate. Which of these two confidence intervals are wider. Why?
   (h) Give a prediction for the hardness that you would expect for an individual piece of plastic after 43 hours; give a 95% prediction interval for this quantity. Is the interval in h) wider than the interval in g)? Why does this always happen?
   (i) To calculate the power for the slope, assume n = 16, σ^2 = MSE and SS_{XX} = 1280. (Note: you can obtain the last value via SAS by using

```
proc univariate data = (dataset name);
var time;
run;
```

   and looking at the output titled "`Corrected SS`" in the `Moments` section.)
      i. Find the power for rejecting the null hypothesis that the regression slope is zero using an α = 0.05 significance test when the alternative is β_1 = 0.5.
      ii. Plot the power as a function of β_1 for the values of β_1 between -1 and +1 in increments of 0.01.

2. (3 pts.) An investigative study collected 16 observations from the Wabash river at random locations near Lafayette. Each observation consisted of the concentration of particulates (X) and fish count (Y). The researchers are interested in how the amount of particulates of the water affects the number of fish. Complete the following ANOVA table for the regression analysis (the p-value need not be exact and is BONUS). Please show your work for the calculation of all of the values. State the null and alternative hypotheses for the F-test as well as your conclusion in sentence form. (Hint for calculating a P value for the F test statistic (BONUS): Use Table B.4, using the numerator df and denominator df, bracket your F value and report the bracketed values for 'A'.)

Source	degrees of freedom	Sum of Squares	Mean Square	F-value	Pr > F
Model	_____	4.50	_____	_____	_____
Error	_____	_____	_____		
Corrected Total	_____	85.20			

3. (0.5) KNNL Problem 2.3.
4. (0.5 pts.) KNNL Problem 2.18
5. (0.5 pts.) KNNL Problem 2.21. Why? If the answer is yes, show the appropriate equations. If the answer is no, provide a general situation (without numbers) where it is false.
6. (0.5 pts.) Given that $R^2 = SSM/SST$, it can be shown that $R^2/(1 - R^2) = SSM/SSE$. If you have $n = 25$ cases and $R^2 = 0.6$, what is the F statistic value for the test that the slope is equal to 0.