Response Surface Methodology, Stat 579 | Homework 2 | Name: ______________________________

Fall 2014 | Prof. Erik B. Erhardt

Part I. (120 points) I recommend reading through all the parts of the HW (with my adjustments) before starting; this may save you some work.
MMA-RSM Chapter 2: 2.6, 2.12, 2.15, 2.16, 2.20, 2.25.

- Use externally studentized residuals in all residual plots.
- In 2.6 and 2.12 conduct lack-of-fit tests for both models.
- In 2.15 and 2.16 obtain a 95% prediction interval at the indicated values in part (b) also.
- In 2.25 fit the model to the original variables first. Then center the predictors before forming product and square terms and refit the model. What changes?

General: Try to do all calculations in R. All R code for the assignment should be included with the part of the problem it addresses (for code and output use a fixed-width font, such as Courier). Code is used to calculate result; text is used to report and interpret results – do not report or interpret results in the code.

(15pts) **1. 2.6** Heat treating is often used to carburize metal parts, such as gears. The thickness of the carburized layer is considered an important feature of the gear, and it contributes to the overall reliability of the part. Because of the critical nature of this feature, two different lab tests are performed on each furnace load. One test is run on a sample pin that accompanies each load. The other test is a destructive test, where an actual part is cross-sectioned. This test involved running a carbon analysis on the surface of both the gear pitch (top of the gear tooth) and the gear root (between the gear teeth). The data in Table E2.4 are the results of the pitch carbon analysis test for 32 parts.

(a) (5 pts) Fit a linear regression model relating the results of the pitch carbon analysis test (PITCH) to the five regressor variables.

(b) (5 pts) Test for significance of regression. Use $\alpha = 0.05$.

(c) (5 pts) Conduct a lack-of-fit test.

(35pts) **2. 2.12** Exercise 2.6 presents data on heat treating gears.

(a) (5 pts) Estimate σ^2 for the model.

(b) (5 pts) Find the standard errors of the regression coefficients.

(c) (5 pts) Evaluate the contribution of each regressor to the model using the t-test with $\alpha = 0.05$.

(d) (5 pts) Fit a new model to the response PITCH using new regressors x_1 = SOAKTIME × SOAKPCT and x_2 = DIFFTIME × DIFFPCT.

(e) (5 pts) Test the model in part (d) for significance of regression using $\alpha = 0.05$. Also calculate the t-test for each regressor and draw conclusions.

(f) (5 pts) Estimate σ^2 for the model from part (d), and compare this with the estimate of σ^2 obtained in part (b) above. Which estimate is smaller? Does this offer any insight regarding which model might be preferable?

(g) (5 pts) Conduct a lack-of-fit test.

(15pts) **3. 2.15** Consider the heat-treating data from Exercise 2.6.

(a) (5 pts) Find 95% confidence intervals on the regression coefficients.

(b) (10 pts) Find a 95% interval on mean PITCH on TEMP= 1650, SOAKTIME = 1.00, SOAKPCT = 1.10, DIFFTIME = 1.00, and DIFFPCT = 0.80. Also, find a 95% prediction interval.

(10pts) **4. 2.16** Reconsider the heat treating in Exercise 2.6 and 2.12, where we fit a model to PITCH using regressors x_1 = SOAKTIME × SOAKPCT and x_2 = DIFFTIME × DIFFPCT.

(a) (5 pts) Using the model with regressors x_1 and x_2, find a 95% confidence interval on mean PITCH when SOAKTIME = 1.00, SOAKPCT = 1.10, DIFFTIME = 1.00, and DIFFPCT = 0.80.

(b) (5 pts) Compare the length of this confidence interval with the length of the confidence interval on mean PITCH at the same point from Exercise 2.15 part (b), where an additive model in SOAKTIME, SOAKPCT, DIFFTIME, and DIFFPCT was used. Which confidence interval is shorter? Does this tell you anything about which model is preferable?
Repeat with the prediction interval.

(15[pts]) **5. 2.20** In Exercise 2.12 we fitted a model to the response PITCH in the heat treating data of Exercise 2.6 using new regressors x_1 = SOAKTIME × SOAKPCT and x_2 = DIFFTIME × DIFFPCT.

(a) (5 pts) Calculate the R^2 for this model, and compare it with the value of R^2 from the original model in Exercise 2.6. Does this provide some information about which model is preferable?

(b) (5 pts) Plot the residuals from this model versus y and on a normal probability scale. Comment on model adequacy.

(c) (5 pts) Find the values of Cook's distance measure. Are any observations unusually influential?

(30[pts]) **6. 2.25**

An article in the Journal of Pharmaceuticals Sciences (vol. 80, 1991, pp. 971–977) presents data on the observed mole fraction solubility of a solute at a constant temperature to the dispersion, dipolar, and hydrogen bonding Hansen partial solubility parameters. The data are in Table E2.5, where y is the negative logarithm of the mole fraction solubility, x_1 is the dispersion Hansen partial solubility, x_2 is the dipolar partial solubility, and x_3 is the hydrogen bonding partial solubility.

(a) (15 pts) Fit the full second-order model with two-way interactions of y vs x_1, x_2, and x_3.
Fit the model to the original variables first. Then center the predictors before forming product and square terms and refit the model. What changes?

(b) (5 pts) Test for significance of regression, using $\alpha = 0.05$.

(c) (5 pts) Plot the residuals, and comment on model adequacy.

(d) (5 pts) Use the extra sum of squares method to test the contribution of the second-order terms, using $\alpha = 0.05$.