Homework 4 – Model Adequacy
MATH 739/839

1. **(17 total pts) The data set NCAAFootball2005.JMP contains the statistics for 117 Division I college football teams from 2005. There are 18 potential regressors in the data set and the response is winning percentage, Win Pct. The regressor column titles will provide some clues as to the nature of the regressor for the football aficionados. The column School is not of value for modeling and is not be used in a model. It may be of interest for identifying interesting points, but otherwise it can be disregarded. However, an understanding of American football is not required to complete this assignment successfully.**

   (a) (3 pts.) Fit and state a full MLR model using the Fit Model platform. Discuss the Actual by Predicted plot and your conclusions about the adequacy of the model based upon that plot.

   (b) (3 pts.) For the full model, generate the residual versus predicted plot and comment on any nonrandom patterns that might appear. If you do note any unusual patterns, then comment on the possible cause of the pattern or patterns.

   (c) (2 pts.) Generate the VIF values for the full model. Is there evidence of significant multicollinearity? Explain.

   (e) (3 pts.) Save the studentized residuals to the data table. Use the distribution platform to create a Normal Quantile plot of the residuals. Based on the plot is the normal assumption for the response valid? If you do note any additional unusual patterns, comment on the possible cause of the pattern or patterns.

   (f) (2 pts.) Create a plot of PRESS residuals vs. Predicted values. Do you see any patterns that were not apparent in the ordinary residual plot? Explain.

   (g) (2 pts.) Create plot of the RStudent residuals and comment on the possibilities of outliers in the data?

   (h) (2 pts.) Why is preferable to use the RStudent residuals to evaluate outliers compared to the standardized residuals?

2. **(21 pts) Use the TireTreadDOE data set for this problem. There are three potential regressors (x1 = hydrated silica level, x2 = silane coupling agent level and x3 = sulfur level) and the response Y is a**

measure of abrasion wear on tire treads. The data table contains the results of a designed experiment to study abrasion as a function of the three regressors or factors in the experiment.

(a) (3 pts.) Using Fit Model, fit and state a full MLR model for the response. Do not include any cross product (interaction) or power terms. Based on the actual by predicted plot is there any evidence of lack of fit in the data? Explain.

(b) (2 pts.) Generate the residual by predicted plot, if it did not appear by default. Do you see any evidence of lack of fit in the residuals? Explain.

(c) (2 pts.) Examine the LOF report. Does the test indicate significant lack of fit? Explain.

(d) (2 pts.) Examine the Effect Leverage Plots. Based on the plots does each of the regressors appear significant? Explain. Do the leverage plots provide some insight into potential lack of fit? Explain.

(e) (2 pts.) Generate the PRESS statistic. Compare the PRESS statistic to the SSE in the ANOVA table. Explain why PRESS is larger. What does this say about the prediction capability of this model?

(f) (2 pts.) Now, fit a full factorial model to the data. This should include the main effects, all two-way interactions, and the three-way interaction. Examine the actual by predicted plot and report if any lack of fit is present for this model? Explain.

(g) (2 pts.) Examine the LOF report. Based on the report does significant LOF still exist? Explain

(h) (2 pts.) Examine the Effect Leverage plots for the three main effects. Do these plots give any idea as to what additional terms the model may need to remove the LOF? It is not always possible to fit needed additional terms with the current data set.

(i) (2 pts.) Generate the PRESS report for this model. Compare the PRESS for the full factorial model to the PRESS for the original model in part (a). Which model is preferred if the goal is to develop a prediction model for future batches of tire compound? Explain.

(j) (2 pts.) Generate the VIF values for the full factorial model. Explain why the VIF's have the observed values you see for this data set.