STAT 512 HW3

See Canvas calendar for due date.
36 points total, 2 points per question unless otherwise noted.

Questions 1 through 6 (Grain Yield): Data relating grain Yield (Y) to the number of Days (X) after flowering that harvesting took place was examined in "Determination of Biological Maturity and Effect of Harvesting and Drying Conditions on Milling Quality of Paddy" (J of Ag Engr. Research (1975):353-361.) The data is available from Canvas as "Grain.csv".

Notes:

- For consistency, please use the I() or poly(, raw = TRUE) functions for fitting the quadratic and cubic models.
- For questions 2-4, you do NOT need to include the diagnostic plots in your assignment. Just discuss your findings.

1. Create a scatterplot of Yield vs Days. Include this plot in your assignment.
2. Fit a linear regression model of Yield on Days. Include the parameter estimate information ("Coefficients" table) in your assignment. Examine a plot of the residuals versus predicted values. What does the residual plot suggest? (**4 pts**)
3. Fit a quadratic regression model (including both linear and quadratic terms). Include the parameter estimate information ("Coefficients" table) in your assignment. Examine a plot of the residuals versus predicted values and comment. (**4 pts**)
4. Fit a cubic regression model (including linear, quadratic, and cubic terms). Include the parameter estimate information ("Coefficients" table) in your assignment. Again examine a plot of the residuals versus predicted values and comment. (**4 pts**)
5. In the cubic model (#4), test the hypothesis that the linear, quadratic and cubic regression coefficients are all simultaneously zero. Give the F-statistic and p-value and make a conclusion about the test.
6. Which model would you choose: linear, quadratic or cubic? Justify your choice. Hint: Think about the simplest model that satisfies assumptions.

Questions 7 through 11 (Drug Test): Data was collected to compare 3 drug treatments for leprosy. Three variables are included in the data set:

Drug: drug treatment (A, D or F)
PreTreatment: a pretreatment score of leprosy bacilli
PostTreatment: a posttreatment score of leprosy bacilli

Ten patients were randomly assigned to each Drug treatment (for a total of n = 30 subjects). The goal of the study is to compare mean PostTreatment values for the 3 drugs. Hence, PostTreatment is the response variable and Drug is the predictor variable. But the researchers would like to consider including the PreTreatment value as a covariate. The data is available from Canvas as "DrugTest.csv".

7. Construct side-by-side boxplots of (1) PreTreatment vs Drug and (2) PostTreatment vs Drug. Also construct (3) a scatterplot of PostTreatment vs PreTreatment for all Drugs on

the same plot. Overlay a fitted regression line for each Drug. Include these plots in your assignment. (**4** pts)

8. Fit the one-way ANOVA model (using Drug as the only predictor). Include the ANOVA table and Tukey adjusted pairwise comparisons in your assignment. What can we conclude about differences between the Drugs? (**4 pts**)
9. Now fit the ANCOVA model with NO Interaction (using Drug and PreTreatment as the predictors). Include the ANOVA table and Tukey adjusted pairwise comparisons in your assignment. What can we conclude about differences between the Drugs? (**4 pts**)
10. Comparing your conclusions from #8 (one-way ANOVA) vs #9 (ANCOVA) you should have found different conclusions regarding significant differences between the Drug treatments. Give a brief explanation of why the conclusions change when we include PreTreatment as a covariate. Hint: Consider the boxplots from #7.
11. An alternative approach to the ANCOVA above is to calculate the difference (Diff = PostTreatment – PreTreatment) and use this Diff as the response in a one-way ANOVA model. Do this and include the ANOVA table and Tukey adjusted pairwise comparisons in your assignment. What can we conclude about differences between the Drugs? (**4 pts**)