Statistics 201 Homework
Assigned Mon, Oct 28 and Due Wed, Nov 6

For this assignment use the data set UCD.txt (linked to the website under data sets and in the list of assignments). The file ***does*** include a row with variable names, and the columns are separated with tabs. The data set, collected on 173 UC Davis students (self-reported), includes the following variables:

ID = numbers from 1 to 173, the ID for that student and the row number with the data
alcohol = average number of alcoholic drinks consumed per week
exercise = average hours per week the student exercises
height = the student's height (in inches)
male = indicator variable, 1 if male and 0 if female
dadht = the student's father's height
momht = the student's mother's height

For this assignment, we will use only the last 4 variables, but retain the file for later use with other variables. Note that there are some missing values, designated NA, but cases with NA for the variables used in this assignment have been removed, so some ID numbers are missing.

1. Fit the model predicting height using male, dadht and momht. (This is the model we went over in discussion.) *Call this model Full.* Show the summary. Write a sentence interpreting the coefficient for "dadht" for this model.

2. Fit the model predicting height using male and momht only. *Call this model MomOnly.* Show the summary. Test the null hypothesis that "momht" is useful in this model. Show the hypotheses, the *p*-value, and a conclusion in the context of the situation.

3. Fit the model predicting height using male and dadht only. *Call this model DadOnly.* Show the summary. Write a sentence interpreting the coefficient for "male" for this model.

4. Create two diagnostic plots for the *DadOnly* model and comment on them.

5. Plot the residuals from the *DadOnly* model versus "momht." Include the least squares line. What do you learn from this graph?

6. Compare the three models (*Full, MomOnly, DadOnly*) using two summary measures discussed in class. Which model is best? Which model is second best? Explain.

7. Using the *Full* model, find a 95% confidence interval for the coefficient for "male" and write a sentence interpreting the interval in the context of this situation.