STAT 614 - HW 8 – LAST ONE!!!!

Due: Thursday, December 3, 2020 in Blackboard by 11:59pm.

Instructions: Please type your solutions in a separate document and upload the document in Blackboard as a pdf. I will not be collecting syntax for this assignment. You will need concepts from Chapters 9 through 12 on the multiple linear regression model and the results from HW 7.

Forced expiratory volume (FEV) is an index of pulmonary function that measures the volume of air expelled after 1 second of constant effort. The data set FEV.csv in Blackboard contains determinations of FEV for 654 children ages 3 through 19 who were seen in the Childhood Respiratory Disease (CRD) Study in East Boston, Massachusetts. These data are part of a longitudinal study to follow the change in pulmonary function over time in children. Variables in the data set are the participant ID number, Age (in years), FEV (in liters), Height (in inches), a binary Sex indicator (0 = female/1 = male), and Smoking status (0 = non-smoker/1 = current smoker).

Consider all variables, Age, Height, Sex, and Smoking status, simultaneously in a multiple regression model.

1. Assess the assumptions of the model and make any adjustments. Be sure to look at *all* residual plots. Make any necessary adjustments.
2. Use the diagnostic tools to identify potential influential observations. Which observations are flagged as potentially being influential? (Note: you'll deal with these in 7 below).
3. Is there evidence of a regression effect? Write the appropriate null and alternative hypotheses. Give the test statistic, p-value, and conclusions of the test.
4. Give and interpret the coefficient of determination (R^2). What is the Adjusted-R^2 value?
5. Which of the four explanatory variables have "significant" associations with FEV, after adjusting for the other variables?
6. Find and interpret the 95% confidence interval for the coefficient of Smoking status. This is the adjusted estimate (adjusting for Age, Height, and Sex). How does this adjusted estimate and CI compare to the unadjusted analysis from HW7?
7. Temporarily hold out any outliers you identified in (2). Do any of the results in (3) – (6) change when holding out the outliers? That is, were the outliers influencing the conclusions? If so, discuss the differences.