***E370, Spring 2016***

***Lab Activities week of 04/25/2016***

***Valued at 25 points***

Solve the following problems. Please answer the questions as directed to obtain full credit.

1. Recall the linear model of the number of “friends” using the data set ***friends.xlsx*** from the previous lab activities. We now try to assess the validity of the model by checking if the assumptions of a linear regression are satisfied. (“Female” variable for Gender, where female=1 if the person is female and female=0 if the person is male)
   1. Is there a multicollinearity problem in this model? Explain. (3 points)
   2. Describe how you would test for normality of errors. Sketch your evidence. Do you think this assumption is satisfied in our model? (3 points)
   3. Test the residuals for heteroscedasticity. Sketch the graph below to support your conclusion. (3 points)
   4. Are there any outliers in the data set? How do you know? (3 points)
2. Data set **Xr18-10.xls** contains data on men’s longevity in years, their parents’ and grandparents’ longevity (average age of grandfathers and average age of grandmothers), and their smoking habits (1=yes, 0=no). We use a linear regression model to predict a person’s longevity using all other variables.
   1. Is there any evidence that would lead us to a conclusion that we face a multicollinearity problem in our model? Describe the key evidence that supports your response. (3 points)
   2. Test the residuals for normality. Sketch the evidence to justify the conclusion. Are the errors centered at 0? (3 points)
   3. Test the residuals for heteroscedasticity. Sketch the graph to support the conclusion. (3 points)
   4. Are there any outliers in the data set? How do you know? Comparing the regression coefficients with and without these outliners. Do you think those outliers are considered “influential” outliers? (4 points)