**2019 SAS Classwork (Individual assignment)**

Q2. The file “diamond.dat” has data on the cut, color, clarity, carat and prices of diamonds in US dollars.

Cut is classified as: Fair, Good, Verygood, and Ideal (These levels are in order of quality of cut with ‘Ideal’ being the best and ‘Fair’ being the worst)

Color: D and E (D is better than E)

Clarity is classified as: VVS1, VVS2, VS1 and VS2 (These levels are in order of clarity with VVS1 being the best and VS2 the worst)

We are interested in finding out using dummy variable regressions how to price different combinations of attributes of diamonds.

Using price as a dependent variable, run a regression model to answer the following questions. Note that you have to create dummy variables for cut, color and clarity before you can do the regression.

1. Use a t-test and test if there is a significant difference in prices between color “D” “E” and “F”?
2. Now run a regression model with price as the dependent variable and cut, color, clarity, and carat after coding the dummy variables appropriately. Using the results of the regression model answer the following questions:

1. How much more price would an ‘Excellent’ cut get over a “VeryGood” cut diamond?
2. How much more price would a VVS clarity get over a VS1 clarity diamond?
3. Which of the explanatory variables are significant?
4. Comment on the model fit.
5. What should be the predicted price for a 0.75 carat diamond, color D, ‘VeryGood’ Cut, ’VVS’ clarity?
6. Using Chi-square test, test whether there is a relationship between color and clarity?
7. Using a t-test, test whether the average price is different for symmetry “VG” compared to symmetry “X”?
8. Check using a regression model whether adding symmetry, polish and fluorescence to the above regression model improves the model fit? Comment on the results of this new model.