**Executive Summary**

This summary is to report on which model is the best to be used in predicting the price of diamonds. This summary will include: what benefits this analysis has and what kind of audience would be interested in this, what variables where used and the process, and any conclusions from the results.

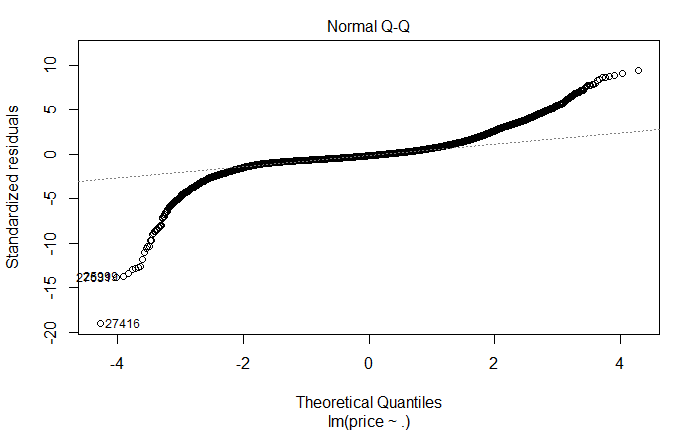
**Practical Use and Target Audience**

The practical use of the Diamond dataset is to see how to calculate a diamond’s price given 9 predictor variables. The target audience would be jewelry stores that want to sell jewelry with diamonds or people who determine if the diamonds are good for jewelry. Predicting a diamond’s price is a very hard process and having a model would allow to help them to get an accurate price.

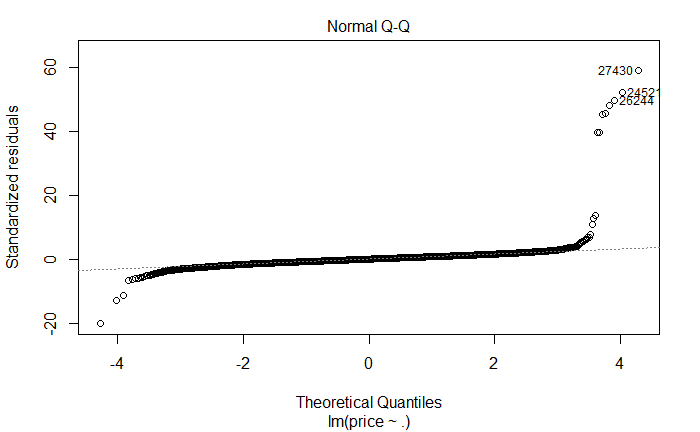
**Variables and Process**

The response variable is price ($326-$18,823), and the predictor variables are carat (), cut, color, clarity, depth (depth percentage), table, x (length in mm), y (width in mm), and z (depth in mm). Some variables that relate to one another are depth, x, y, and z. They are related because they are quantitative variables, and the depth is calculated from x, y, and z. The formula is . The other two quantitative variables are carat and table, and the rest (cut, color, and clarity) are qualitative. Some things I should mention are that the variable, color, means how colorless the diamond is, and not the actual color.

The two predictive modeling processes I chose were Multiple Regression with best subsets and Robust Regression. To come to that conclusion, I used exploratory analysis. The first thing I did was to use a linear regression fit to look at the graphs using the plot() function. When fitting the whole model unchanged, the Normal Q-Q plot looked like this:



This was not as normal as I would like it to be, but I did notice some outliers R picked out. The next thing I did was to transform the response variable by log. The result Normal Q-Q plot now looked like this:



This looked much better to be normally distributed but there seems to be outliers again. So, by concluding that it is much closer to be normal and there are outliers, then Robust Regression with a transformed price would be best as it puts less weight on outliers. Then I tried to reduce the number of predictor variables by using best subsets. When I did the process, I found that no variables were omitted but since the Normal Q-Q was really close to be normally distributed, I decided to test out 3 full models using multiple regression, and robust regression with tukey and huber methods.

Once I figured the two methods, I used Double-Cross Validation to figure out which model is the best and how effective the modeling process was.

**Conclusion**

In conclusion the best model would be model 2 from the selection process. Model 2 would be a full model using Tukey’s bi-square method. The model would look something like this: . By taking this equation you can predict or at least have an estimate what the price of a diamond would be. To use this equation, you would first determine what cut, color and clarity would be and put a one in the corresponding category and a zero in the rest. For the other quantitative variables, you would determine their value and put them in the equation. Finally you would have to take that number and put it into the “x” slot in to cancel out the log to get the full price.

Some interesting observations I saw from the final model were from the cut variable. The numbers would make sense as cutIdeal has the highest multiplier at .158 since that is the best kind of cut you can have. Another observation I noticed was that color with the lowest number of -.518 belongs to J which is considered the worst. It makes sense because if the diamond has colorJ, then the price will decrease a lot.

Since the R-squared is 0.96 for the double-cross validation, this means that the overall assessment of how well the modeling process will predict new observations is 96%. This is extremely good as it is very close to 100%.

**Reference**

B. (n.d.). Diamond Color. Retrieved December 16, 2020, from https://www.bluenile.com/education/diamonds/color

**Variables**

