# MAT 303 Module Two Problem Set Report

Interaction Terms and Qualitative Predictors

Camilo Hoyos

camilo.hoyos@snhu.edu

Southern New Hampshire University

## 1. Introduction

The mtcars dataset contains attributes of certain cars. Attributes include make and model, miles per gallon, cylinder, disposition, horsepower, weight, and rear axle ratio. The analysis that I will be performing is identifying the relationship between fuel economy, the response variable, and weight, horsepower, and rear axle ratio the predictor variables. The results can be used to predict a vehicle’s fuel economy based on weight, horsepower, and rear axle ratio. It would be great to predict a prototype and what their fuel economy will be before it is manufactured.

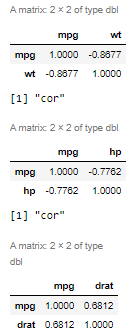
## 2. Data Preparation

The fuel efficiency (mpg), weight (wt), gross horsepower (hp), and rear axle ratio (drat) are the important variables in this data set. These variables are what will be used to determine relationships and a multiple regression model with an interaction term. There are a total of 12 columns and 32 rows.

## 3. Model with Interaction Term

### Correlation Analysis

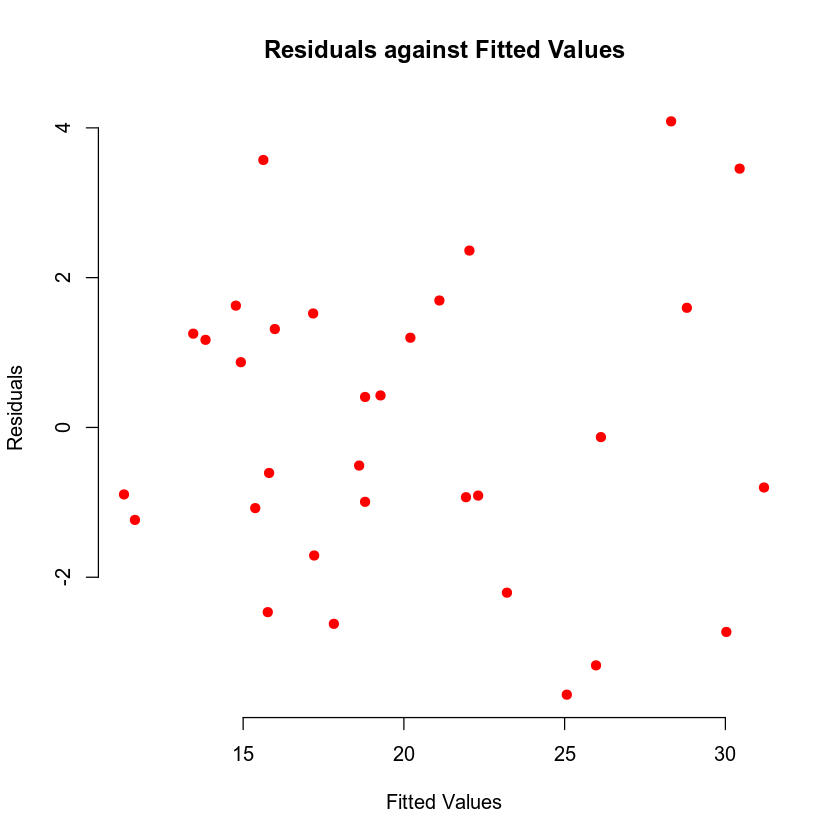
There is a Pearson correlation coefficient(PCC) calculated for 4 pairs of variables. Starting with fuel economy and weight we see a PCC of -0.8677. What this means is that as weight increases, fuel efficiency is expected to drop every 0.8677 units of weight. Between fuel economy and horsepower we see a PCC of -0.7762. This describes that as horsepower increases, fuel efficiency is expected to drop every 0.7762 units of weight. Lastly, fuel economy and rear axle ratio we see a PCC of 0.6812. This explains that as the rear axle ratio increases, fuel efficiency is expected to rise every 0.6812 units.



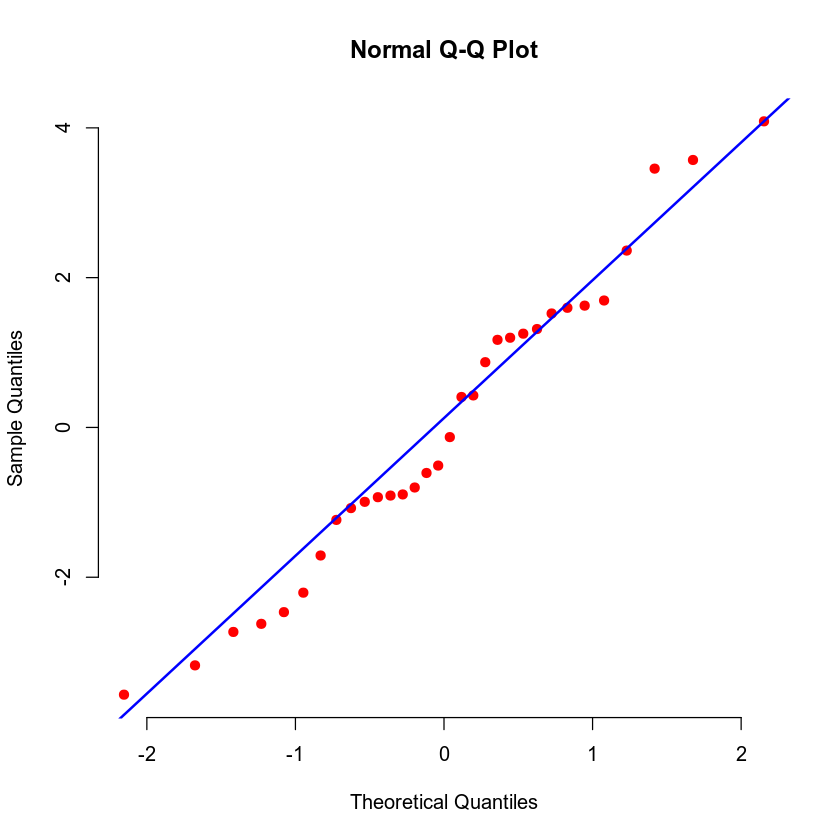
### Reporting Results

The general form of the regression model for fuel economy, weight, horsepower, and rear axle ratio with interaction terms looks like: , where X1 is weight, X2 is horsepower, and X3 is rear axle ratio. X1X2 is the interaction term for weight and horsepower, and X1X3 is the interaction term for weight and rear axle ratio. The multiple regression model for fuel efficiency is . The coefficient of multiple determination is 0.8907 or 89.07%. What this describes is that 89.07% of the variability of fuel economy (mpg) can be predicted from weight, horsepower, and rear axle ratio. It should be noted that our previous multiple regression model’s coefficient of multiple determination was 82.68%, this means it has more predictability; however, falls closer to multicollinearity since some of these variables correlate. The adjusted coefficient of multiple determination is 0.8697 or 86.97%. Adjusted coefficient of multiple determination provides how well the model fits the sample data being used. If we assume the weight of a vehicle is 3.5 (3500 lbs) our model becomes which simplifies to . So with a weight of 3.5, each unit of horsepower (X2) will decrease fuel efficiency by 0.022385. Likewise, with a weight of 3.5, each unit of rear axle ratio (X3) will increase fuel efficiency by 0.5228. By having one of the variables, we are able to simplify all coefficients with X1 and simplify the model to two variables. Now the new coefficients are the relationships of the leftover variables that allow us to identify the strength of these variables related to the provided weight.

Below you will see the residuals against fitted values.

**

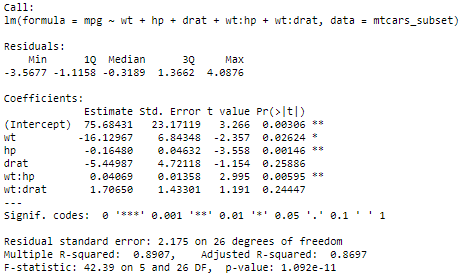
Below you will see the normal Q-Q plot.

**

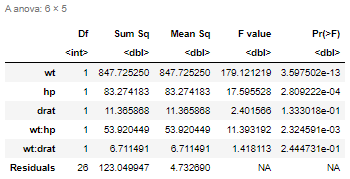
From the residuals against fitted values plot we can identify if there is no nonlinear dataset. This ensures we do not violate the Constant Variance assumption. The normal Q-Q plot assists with the normality assumption. If the plotted dots fit snugly on the line, we can assume that the assumption of Normality holds.

### Evaluating Model Significance

The null hypothesis (H0) is that there is no relationship between fuel efficiency, weight, horsepower, and rear axle ratio. The alternative hypothesis (Ha) is that there is a relationship between fuel efficiency, weight, horsepower, and rear axle ratio where the fuel efficiency can be predicted. From the F-test we can see that the P-value is 1.092e-11. This is lower than the 5% level of significance identified in the report. Seeing that the P-value is lower than the level of significance we can reject the null hypothesis (H0).

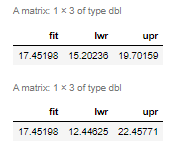
**

Utilizing the analysis of variance (ANOVA) we can see the P-value of each predictor variable. Each predictor variable, with the exception of rear axle weight and the interaction term of weight and rear axle weight, is lower than 5% or 0.05. Weight is at 3.5e-13, horsepower is at .00028, and the interaction term of weight and horsepower is at .00232. Both the rear axle weight, 0.133, and the interaction term of weight and rear axle weight, 0.244 are above the significance level. Regarding rear axle weight, this identifies that these particular predictor variables would not be capable of rejecting their null hypothesis (Ho). The null hypothesis for these would be that there is a relationship between fuel economy and rear axle ratio. Both the predictor variable and the interaction term fail the level of significance test. With that, it would be recommended to remove the rear axle weight predictor variable and interaction term.

**

### Making Predictions Using the Model

The predicted fuel economy of a car that weighs 2.965 (2965 lbs) units, has 210 gross horsepower, and a rear-axle ratio of 2.91 would be 17.45198 miles per gallon. The 95% prediction interval has a fit of 17.45198, a lower limit of 12.44625 and an upper limit of 22.45771. The 95% confidence interval has a fit of 17.45198, a lower limit of 15.20236 and an upper limit of 19.70159. This provides that there is a 95% probability the sample will land in the lower and upper limit. The distinction to take into account is that the prediction interval takes uncertainty into account.

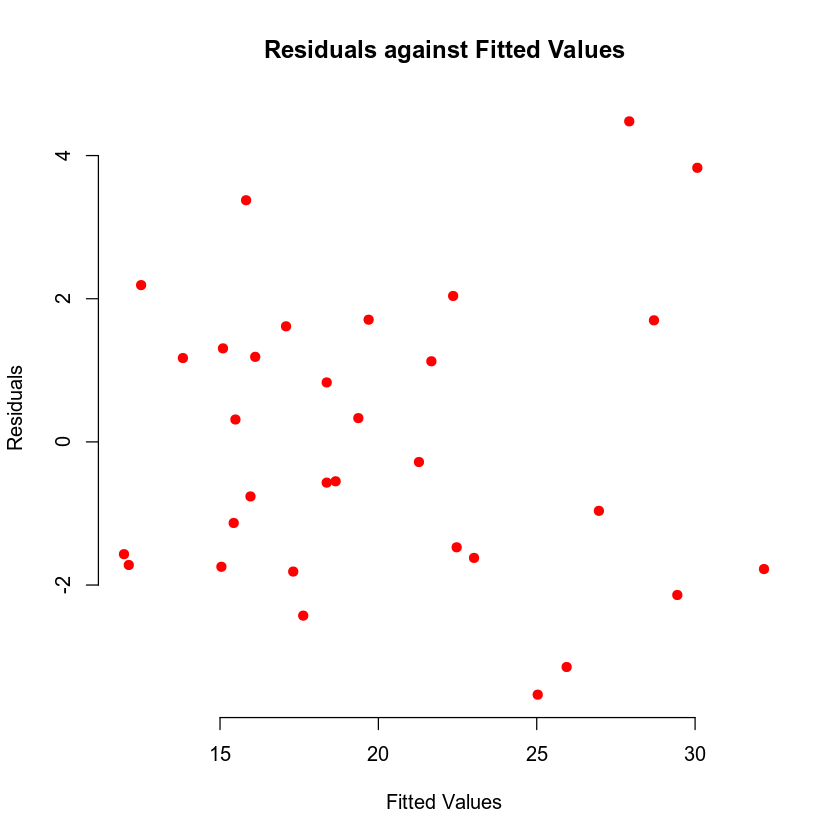
**

## 4. Model with Interaction Term and Qualitative Predictor

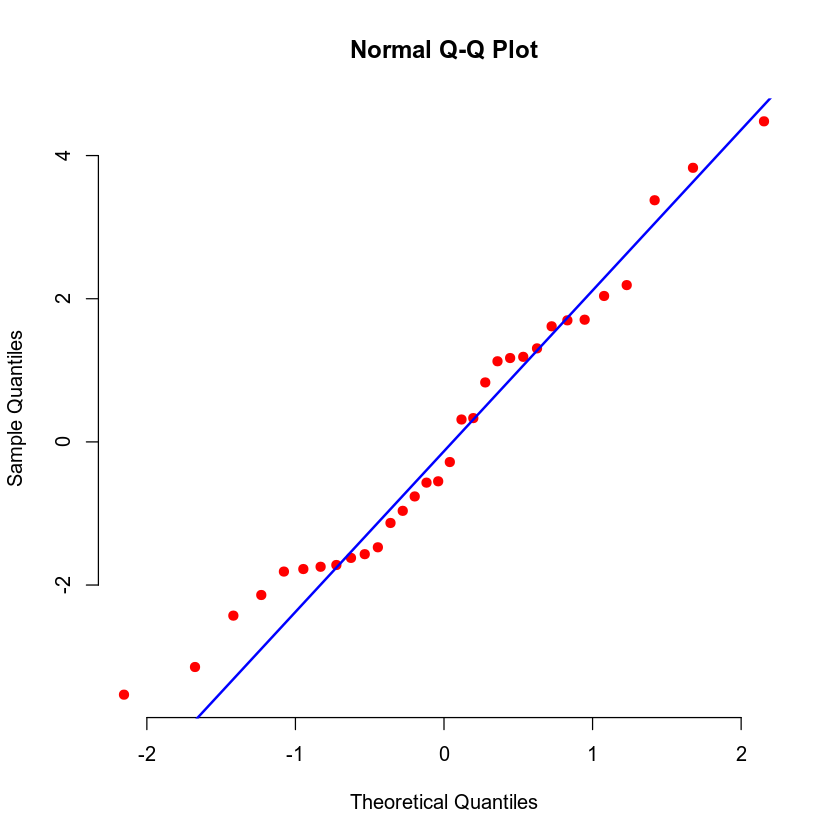
### Reporting Results

The multiple regression model is where Y= miles per gallon, X1 = weight, X2 = horsepower, X1X2 is the interaction term for weight and horsepower, X3 = 6 cylinders, and X4 = 8 cylinders. The multiple regression model for fuel economy is . The coefficient of multiple determination is 0.888 or 88.8%. This states that 88.8% of the variance in fuel economy can be related to weight, horsepower, the interaction between weight and horsepower, and the number of cylinders. The adjusted coefficient of multiple determination is .8664 or 86.64%. The coefficient of multiple determination which explains that 88.8% of the variance in fuel efficiency can be explained by weight, horsepower, the interaction between weight and horsepower, and the number of cylinders. The adjusted coefficient of multiple determination simply provides how well the model fits the sample data.

Below you will see the residuals against fitted values.



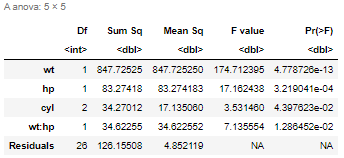
Below you will see the normal Q-Q plot.



We can see from the residuals against fitted values that there is no nonlinear pattern so we do not violate the constant variance assumption. We can confirm this is homoscedasticity. The normal Q-Q plot assists with normality. We see that the plotted points line up closely against the line provided so the normality assumption has not been violated.

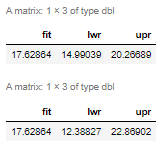
### Evaluating Model Significance

The null hypothesis (H0) is that there is no relationship between fuel economy, weight, horsepower, cylinders, and the interaction between weight and horsepower. The alternative hypothesis (Ha) is that there is a relationship between fuel economy, weight, horsepower, cylinders, and the interaction between weight and horsepower that allows for prediction. All predictor variables have a 5% level of significance. Weight has a p-value of 4.77e-13, horsepower has a p-value of 0.0003, cylinders has a p-value of 0.044, and the interaction of weight and horsepower has a p-value of 0.0129.



### Making Predictions Using the Model

The predicted fuel economy for a car that has a 2.965 (2965 lbs) weight, 210 horsepower, and 6 cylinders is 17.62864 miles per gallon. The 95% prediction interval has a fit of 17.62864, a lower limit of 12.38827, and an upper limit of 22.86902. The 95% confidence interval has a fit of 17.62864, a lower limit of 14.99039, and upper limit of 20.26689. This states that 95% of the sample will land within the upper and lower limit. The prediction interval has a wider range because it takes regression error () into account from the population model.

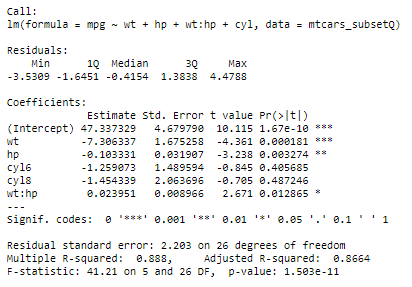


## 5. Conclusion

* *What is the practical importance of the analyses that were performed?*

Based on what variables have been utilized for prediction and the sample size, I would utilize the second multiple regression model : . The first model has some multicollinearity between the interaction terms as we see some p-values that fail the significance test. The p-values of the second multiple regression model are more acceptable.

What the second model provides is a means to predict the fuel efficiency of a vehicle assuming we know the weight (X1), the horsepower (X2), and the number of cylinders (X3 or X4). We have enough evidence using an F-test P-value = 1.503e-11 and that we have an adjusted coefficient of multiple determination that is not as high as our first model, but also not too low that we didn’t find a good model that fits our data. We can reject our null hypothesis with the second model with a high level of confidence.



The practical importance of this analysis is that there is a feasible method to predict the fuel economy of a vehicle. So long as we have the three predictor variables, weight, horsepower, and number of cylinders, we can make a fair prediction of a vehicle’s fuel efficiency. This prediction can assist with decision making; for instance, what if someone wanted to determine from a specification of the three predictor variables, would it hit a target fuel efficiency knowing current technologies? If the answer fails to meet this target fuel efficiency, one knows a new technology or and advancement would be required to achieve the target fuel efficiency.