# MAT 303 Module Two Problem Set Report

Interaction Terms and Qualitative Predictors

Gigi Morales

gissel.morales@snhu.edu

Southern New Hampshire University

## 1. Introduction

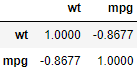
The data set being explored pertains to car maker data specifically associated with fuel economy. The results of this analysis can be used to better estimate what factors negatively or positively impacts a vehicle’s fuel economy. To do so, this analysis will consist of different multiple regression models with a variety of interaction terms and qualitative predictors.

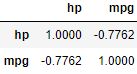
## 2. Data Preparation

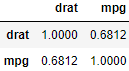
The important variables in this data set are weight, horsepower, rear axle ratio, and number of cylinders. This data set contains 32 rows and 12 columns.

## 3. Model with Interaction Term

### Correlation Analysis







It quickly becomes apparent that, unlike weight and horsepower which both have a strong negative effect on fuel economy, rear axle ratio has a positive effect on fuel economy even though it is not nearly as strong as the negative effect of the aforementioned variables.

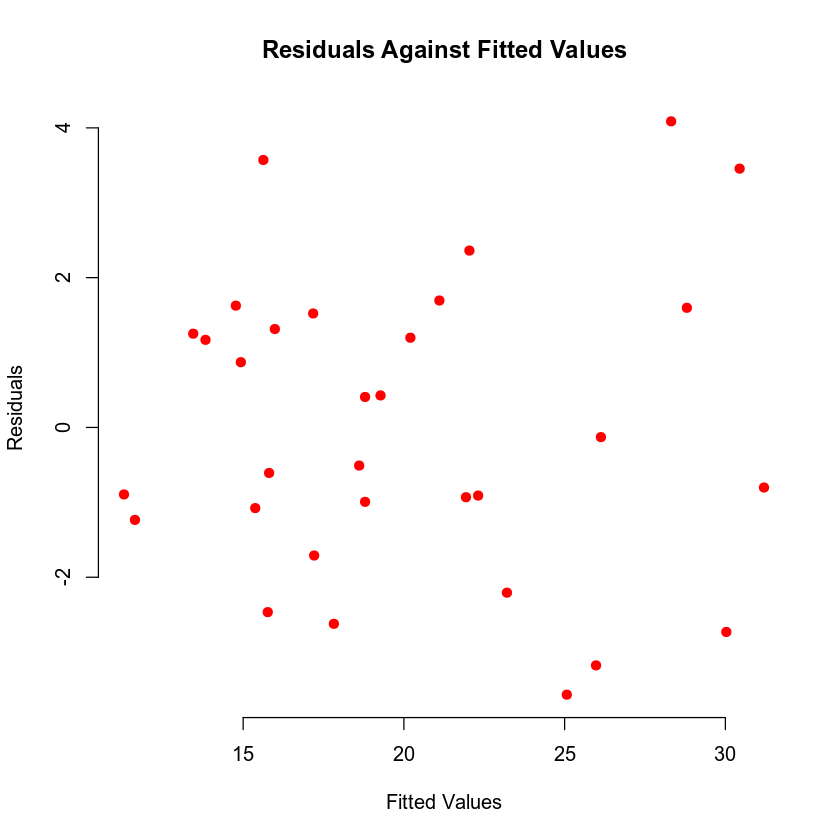
### Reporting Results

The value of *R-Squared* is 0.8907 while the value of *Adjusted R-Squared* is 0.8697. This means that the predictors within this model can account for about 86-89% variance in the fuel economy of a given vehicle. For a car with weight 3.50 the calculation for change in fuel economy per unit increase in horsepower is as follows:

According to this model, the estimated change in fuel economy per unit of increase in horsepower of a vehicle with a weight of 3.50 is -0.307215 units.

According to this model, the estimated change in fuel economy per unit of increase in rear axle ratio for a vehicle with a weight of 3.50 is 0.52288 units.

Chart, line chart, scatter chart

Description automatically generated

First looking at the residual against fitted values plot, there is no discernable pattern lending credence to the assumptions of both linearity and homoscedasticity. Furthermore, upon evaluating the Q-Q plot is also becomes apparent that the assumption of normality also holds true due to the lack of outliers.

### Evaluating Model Significance

The null hypothesis for the overall F-test states that the model is not significant at a 5% level of significance whereas the alternative hypothesis states that the model is significant at the same level of significance. With a significance level of 0.05, when considering the model’s calculated p-value of 1.092e-11 it becomes apparent that the model falls well below the significance. This then means we can reject the null hypothesis in favor of the alternative hypothesis for this model.

The null hypothesis for the individual beta test, much like that overall F-test, states that the different terms within the model are not significant at a level of 5%. On the other hand, the alternative hypothesis states that the terms are significant at the same level of significance. The p-values are as follows:

Weight – 0.02624

Horsepower – 0.00146

Rear Axle Ratio – 0.25886

Wt:Drat – 0.24447

Wt:Hp – 0.00595

With this information it can be concluded that three out of the five different predictors are significant. This means that for weight, horsepower, and the interaction term between weight and horsepower the null hypothesis can be rejected in favor of the alternative hypothesis. However, for rear axle ratio and the interaction term between weight and rear axle ratio the null hypothesis can be accepted due to their failure to fall below the 0.05 level of significance.

### Making Predictions Using the Model

According to this model, the fuel economy for a vehicle with a weight of 2.965, horsepower of 210, and a rear-axle ratio of 2.91 is 17.452. The 95% prediction interval has a lower bound of 12.4462 and an upper bound of 22.4577. This means that the model predicts with 95% certainty that an individual vehicle of the same weight, horsepower, and rear-axle ratio will have a fuel economy that fall within the mentioned bounds. The 95% confidence interval has a lower bound of 15.2024 and an upper bound of 19.7016. This can be interpreted to mean that the model is 95% confident that the average vehicle with the same weight, horsepower, and rear-axle ratio will fall within its mention parameters.

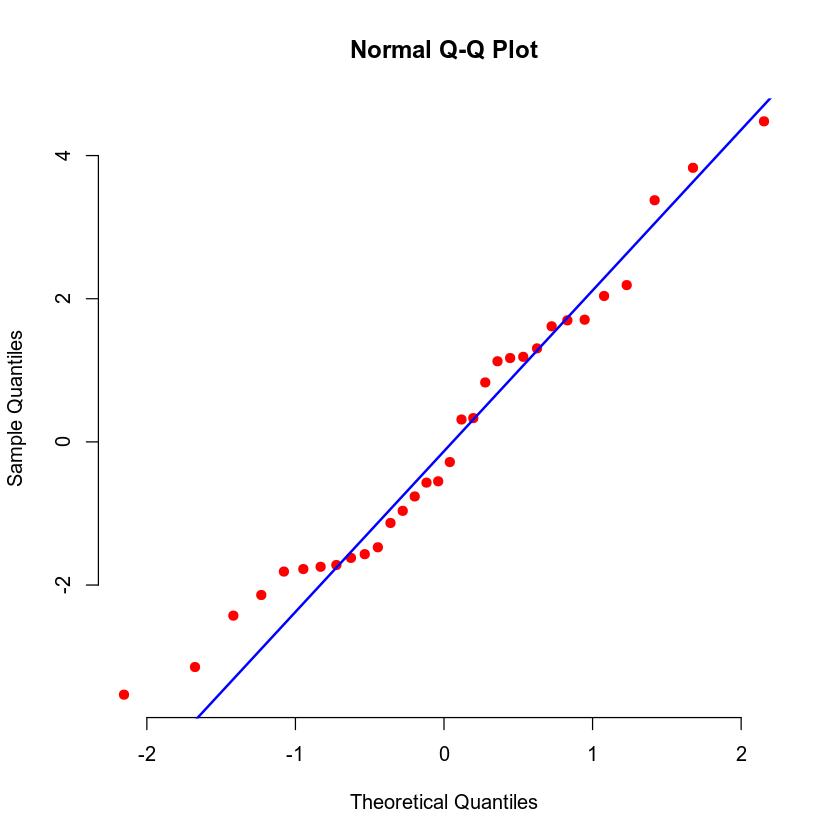
## 4. Model with Interaction Term and Qualitative Predictor

### Reporting Results

The calculated value of R-squared and Adjusted R-squared are 0.888 and 0.8664 respectively. This means that the model can account for about 87-89% of variance in fuel economy with the predictors presented.

Chart, scatter chart

Description automatically generated



When assessing the residuals against fitted plot, there again is no discernable pattern which aids in lending credence to the assumption of homoscedasticity. Furthermore, the Q-Q plot shows no evidence of severe outliers from the regression line which helps the normality assumption hold true.

### Evaluating Model Significance

The null hypothesis for the overall F-test states that the model is not significant at a 5% level of significance. Whereas the alternative hypothesis states that the model is significant at the same confidence level. The model’s calculate p-value is 1.503e-11 falling well below the 0.05 level being tested, meaning that the null hypothesis can be rejected in favor of the alternative hypothesis.

In terms of the different quantitative and qualitative predictors, the null hypothesis for the individual beta test states that the predictors are not significant at a 5% level of significance. The alternative hypothesis, however, conversely states that the predictors are significant at the same confidence level. The p-values of the predictors are as follows:

Wt – 0.000181

Hp –0.003274

Cyl6 – 0.405685

Cyl8 – 0.487246

Wt:Hp – 0.012865

Much like our previous model, three out of the five predictor) are significant within the model while the other two (the qualitative predictors) are not at a 5% level of significance. In this model, the null hypothesis can be rejected for weight, horsepower and the interaction term between weight and horsepower in favor of the alternative hypothesis. However, the null hypothesis can be accepted for the qualitative predictors of Cyl6 and Cyl8 due to falling above the tested level of significance.

*Evaluate model significance for the regression model. Address the following questions in your analysis:*

### Making Predictions Using the Model

The predicted fuel economy for a vehicle with a weight of 2.965, a horsepower of 210, and a has six cylinders is 17.6286 according to this model. The 95% prediction interval has a lower bound of 12.3883 and an upper bound of 22.869. This means that the model is 95% confident that a given vehicle of the same parameters will fall within the bounds of the interval given. The 95% confidence interval has a lower bound of 14.9904 and an upper bound of 20.2669. This means that the model is 95% confident that the average fuel economy for vehicles with the given parameters is within the mentioned bounds. The prediction interval is always wider than the confidence interval because the prediction interval does not account for sampling uncertainty related to estimating the regression parameters.

## 5. Conclusion

Based on the analysis performed and assuming a sufficiently large sample size, I would recommend the second model due to the lower p-values associated with the model as well as having a slightly higher R-squared value. With this information, car manufacturers can better account for negative impacts on fuel economy when manufacturing a new vehicle without severely limiting other factors.