# MAT 303 Module Two Problem Set Report

Interaction Terms and Qualitative Predictors

Ashley De Venuto

ashley.devenuto@snhu.edu

Southern New Hampshire University

## 1. Introduction

I’ve been given a dataset to analyze a car’s fuel economy. As an analysist, I’m here to run statistical analysis to find any association between a vehicle’s fuel efficiency and other variables. For the first model, we want to figure out how horsepower, quarter mile time, and rear axle ratio relate to fuel economy. For the second model, we are looking at how horsepower, quarter mile time, and number of cylinders correlate with fuel efficiency. By using the results of these models, the car manufactures can determine the best fuel efficiency for each vehicle. The manufactures can also see how these variables can affect fuel economy. For this analysis, I will run two different multiple regression models. The first regression model will be comparing three variables and two interaction term variables. The second regression model will be comparing two variables, one interaction term variable, and one qualitative predictor variable. I will also run analysis on fitted values and residuals and will create a visual scatterplot to represent the results. I will also create a Q-Q plot to show the normality assumptions for the residuals. I will also determine what the confidence intervals for each model. I will use the results of the analysis to determine if the variables in each model have a statistically significant influence on fuel economy.

## 2. Data Preparation

I will be looking at certain factors within the data set to complete both multiple regression models. These factors (variables) include horsepower, quarter mile time, rear axle ration and cylinder against miles per gallon. I will use these factors (variables) to determine whether they have a statistically significant impact on fuel economy. There are 32 rows representing the type of vehicles and 12 columns representing the variables of the vehicles within the dataset. I only printed the first 6 rows of the dataset.

## 3. Model with Interaction Term

### Correlation Analysis

A table with numbers and symbols

Description automatically generated  
Here is the results of the Pearson Correlation Coefficients between fuel economy(mpg) and horsepower(hp); fuel economy and quarter mile time(qsec); and fuel economy and rear axle ratio(drat). Horsepower has a negative association with fuel economy with a correlation coefficient of -0.7762. The other two variables have a positive correlation coefficient of 0.4187 for quarter mile time and 0.6812 for rear axle ratio respectfully. The results show that the fuel economy suffers as the amount of horsepower rises. The results also show that as quarter mile time and rear axle ratio go up so does the quality of the fuel economy.

### Reporting Results

The equation below shows the general form of the regression model for fuel economy, horsepower, quarter mile time, rear axle ratio and includes interaction terms for horsepower & quarter mile time and horsepower & rear axle ratio:

Here is the prediction model:

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><mi>y</mi><mo>&#x2227;</mo></mover><mo>=</mo><mo>&#xA0;</mo><mover><msub><mi>&#x3B2;</mi><mn>0</mn></msub><mo>&#x2227;</mo></mover><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>+</mo><mo>&#xA0;</mo><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mo>&#xA0;</mo><mover><msub><mi>&#x3B2;</mi><mn>3</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>3</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>4</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mo>&#xA0;</mo><mover><msub><mi>&#x3B2;</mi><mn>5</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>3</mn></msub></mstyle></math>","origin":"MathType for Microsoft Add-in"}

Here is the regression model for fuel economy using the data set horsepower, quarter mile time, and rear axle ratio as predictors as well as the interaction terms for horsepower and quarter mile time and horsepower and rear axle ratio:

For has a value of 0.8207. This means that horsepower, quarter mile time, and rear axle ratio account for 82% of the whole variation in fuel economy. For has a value of 0.7862. This value represents the modified for the number of predictors in the model.

This model will help calculate the change in fuel economy of a car with 160 horsepower for each unit increase in quarter mile time. The regression equation is:

Looking at the results, it looks like with every unit increase in quarter mile time, the fuel economy of a car will decrease by 1.4824 when the car has 160 horsepower.

This model will help calculate the change in fuel of a car with 160 horsepower for each unit increase in rear axle ration. The regression equation is:

Looking at the results, it looks like with every unit increase in rear axle ratio, the fuel economy of a car will increase by 0.3546 when the car has 160 horsepower.

I have created scatterplots for the residuals against fitted values and the standard Q-Q plot after obtaining the fitted values and residuals using the model for the dataset.

A diagram with red dots

Description automatically generated A graph of a normal q-q plot

Description automatically generated

The data shown in the regression against fitted values scatterplot shows homoscedasticity because there’s no discernible pattern. When looking at the Q-Q plot, it shows that the values are Normal. The residuals are normal due to data on the scatterplot not deviating too much from the line.

### Evaluating Model Significance

While trying to find the regression model’s significance at 5% level of significance, I had to conduct an overall F-test. Here is where we find our null hypothesis and alternative hypothesis. Here are those hypotheses:

After running the model, the p-value = 6.098e-09. This is less than the 5% level of significance. This means we must reject the null hypothesis and accept the alternative hypothesis. This also means that there is at least one predictor variable that has a statistically significant relationship with the fuel efficiency.

To determine which variables in the model are significant at a 5% level of significance, I need to test each variable individually. The null hypothesis and alternative hypothesis are listed below:

Horsepower has a p-value of 0.0118, quarter mile time has a p-value of 0.0404, rear axle ration has a p-value of 0.0326, horsepower versus quarter mile time has a p-value of 0.0031, and horsepower versus rear axle ratio has a p-value of 0.0841. All the variables are significant at 5% except for horsepower versus rear axle ratio. It can be determined that all but hp:drat have an influence on the fuel economy of a car. I can also conclude that horsepower versus rear axle ratio have no statistically significant link between these specific variables and fuel efficiency. P-values are rounded up to the nearest fourth digit.

### Making Predictions Using the Model

To complete the predictions for fuel economy, I will be using multiple regression model. I will use the regression model to predict the fuel economy of a car when it has 175 horsepower, 14.2 quarter mile time, and has a 3.19 rear axle ratio. The multiple regression model will look something like this:

The estimated fuel efficiency for this specific vehicle is 21.5285 mpg. The 95% prediction interval for the fuel economy for this vehicle is (15.0897 – 27.9674). The results show that this range of numbers is where an estimated 95% of fuel economy will land. The 95% confidence interval for the fuel economy of this car is (18.5881 – 24.469). The results show that this range is where 95% of fuel economy would land if we repeatedly sampled cars with these characteristics.

## 4. Model with Interaction Term and Qualitative Predictor

### Reporting Results

I built a second multi regression model with an interaction term and qualitative predictor. The general form and prediction equations for fuel economy using horsepower, quarter mile time, interaction term for horsepower and quarter mile time, and number of cylinders is shown below:

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><mi>y</mi><mo>&#x2227;</mo></mover><mo>=</mo><mover><msub><mi>&#x3B2;</mi><mn>0</mn></msub><mo>&#x2227;</mo></mover><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>3</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>4</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>3</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>5</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>4</mn></msub></mstyle></math>","origin":"MathType for Microsoft Add-in"}

The regression model for fuel economy using horsepower, quarter mile time, interaction term for horsepower and quarter mile time, and number of cylinders is shown below:

The has a value is 0.8327 and the has a value is 0.8005. This means that predictors for horsepower, quarter mile time, number of cylinders, and the interaction term of horsepower versus quarter mile time can account for 83% of the variation in fuel efficiency.

A diagram with red dots

Description automatically generated A graph with blue dots

Description automatically generated

Since there is no discernable pattern, the data then meets the condition for homoscedasticity. This is shown in the scatterplot where you see residuals against fitted values are plotted. The Q-Q plot demonstrates how the quantile originated from Normal distributions.

### Evaluating Model Significance

To see whether the regression model is significant at a 5% level of significance, I had to run an overall F-test. I had to first identify the null hypothesis and the alternative hypothesis. The null and alternative hypothesis are shown below:

The p-value for this model is 2.526e-09. This is less than the 5% level of significance. This means that we should reject the null hypothesis and accept the alternative hypothesis. This also means that al least one of the predictor variables have a statistically significant correlation.

To be able to determine which variables are significant at a 5% level of significance, I must carry out individual tests. I will do this by identifying the null hypothesis, the alternative hypothesis, and the p-value. The null and alternative hypothesis are shown below:

Horsepower has a p-value of 0.0848, quarter mile time has a p-value of 0.4828, cly6 has a p-value of 0.0118, cyl8 has a p-value of 0.0847, and horsepower versus quarter mile time has a p-value of 0.0246. The only variables that a statistically significant at a 5% level of significance is cly6 and the interaction term of horsepower versus quarter mile time. The rest of the p-values are more than the 5% significance level. This does mean that the null hypothesis should be rejected, and the alternative hypothesis should be used. The conclusion of this model is that only cyl6 and the interaction term of horsepower against quarter mile time have a statistically significant link to fuel economy.

### Making Predictions Using the Model

The estimated fuel efficiency for a vehicle that has 175 horsepower, 14.2 quarter mile time, and 6 cylinders is by running a prediction model using the data within model2. According to the model being used, the fuel economy of this vehicle would achieve 21.3424 mpg. The 95% prediction interval for the fuel economy of this car is (14.8764 – 27.8085). This range of numbers is where 95% of the cars fuel economy will fall into if the car has 175 horsepower, 14.2 quarter mile time, and 6 cylinders when interpreting the prediction interval. The 95% confidence interval for the fuel economy for this car is (17.9965 – 24.6884). This can be interpreted by saying that if a collection of cars were with the predicted vales of 175 horsepower, 14.2 quarter mile time, and 6 cylinders, there is a 95% confidence level for the range of numbers. The prediction intervals are wider than the confidence intervals due to the uncertainty of the mean and the random variation of the individual values.

## 5. Conclusion

Based on my analysis, I would recommend using model1. In the first model, horsepower, quarter mile time, rear axle ratio and horsepower versus quarter mile time have a significant influence on the cars fuel economy. The only variable that doesn’t have a significant influence on fuel efficiency is horsepower versus rear axle ratio. The four variables with a strong relationship with fuel economy is also good at predicting milage per gallon. With the results of the analysis out, car manufacturers can use these findings to help build fuel efficient cars in the future. To get the most mileage out of the car, the manufacture should consider having less horsepower, more rear axle ration, and more quarter mile time for the vehicle.