# MAT 303 Module Two Problem Set Report

Interaction Terms and Qualitative Predictors

Brandon B. Gibbs

Brandon.gibbs@snhu.edu

Southern New Hampshire University

## 1. Introduction

This report analyzes the "mtcars" dataset to understand the relationship between various car attributes and fuel economy miles per US gallon(mpg). The primary objective is to develop a regression model that accurately predicts fuel economy based on key predictors horsepower (hp), quarter mile time (qsec), and rear axle ratio (drat), including interactions and qualitative variables.

The results of this analysis have important implications for automotive manufacturers, offering data-driven recommendations for engineering practices that could improve fuel economy. By focusing on reducing horsepower while optimizing rear axle ratios, manufacturers can create more efficient engines that meet regulatory standards and align with consumer demands for better fuel economy.

## 2. Data Preparation

The "mtcars" dataset features 32 car types(rows) and 11 variables(columns) affecting fuel economy (mpg). This analysis focuses on four key variables: Miles per US Gallon (mpg) measures fuel economy, indicating the distance traveled on one gallon of gasoline; Gross Horsepower (hp) represents the engine's theoretical power output and performance; Rear Axle Ratio (drat) indicates the drive shaft's revolutions for each complete rear wheel rotation; and Quarter Mile Time (qsec) reflects the time it takes for a vehicle to complete a quarter mile. Quarter of a mile from a standing start. By analyzing these variables, we aim to understand how engine performance and acceleration characteristics influence overall fuel economy.

## 3. Model with Interaction Term

### Correlation Analysis

There is a strong negative correlation between fuel economy and horsepower (-0.7762), indicating that as horsepower(hp) increases, fuel economy tends to decrease. In contrast, there is a moderate positive correlation between fuel economy and quarter mile time (0.4187), suggesting that cars with better fuel economy may have longer quarter mile time(qsec). Additionally, there is a strong positive correlation between fuel economy and rear axle ratio (0.6812), which means that cars with higher rear axle ratios(drat) tend to have better fuel economy.

### Reporting Results

The general form of this regression model equation is:

The prediction regression model equation is:

*From the output the prediction model equation:*

*ŷ = -14.529137 + 0.352800x₁ + 1.509555x₂ + 5.666624x₃ - 0.018723(x₁x₂) - 0.033246(x₁x₃)*

**R-squared (R²)** value is 0.8207, indicating that approximately 82.07% of the variability in the dependent variable (mpg) can be explained by the independent variables (hp, qsec, drat, hp:qsec, hp:drat) in the model. The **Adjusted R-squared (R²a)** value is 0.7862, showing that about 78.62% of the variability in mpg is explained by the model, adjusted for the number of predictors.

The change in fuel economy (miles per US gallon) for each one-second increase in quarter mile time (qsec) with a 160 horsepower (hp) car is represented by the coefficient for qsec, which is 1.509555. The interaction between horsepower and quarter mile time, has a coefficient of -0.018723. Therefore, for a car with 160 horsepower, you can estimate how fuel economy changes with each additional second of quarter mile time by the following calculation:

Each one-second increase in quarter mile time (qsec) results in a decrease of about 1.49 mpg in fuel economy for a car with 160 horsepower.

When the rear axle ratio (drat) increases by one unit in a car with 160 horsepower (hp), the fuel economy (mpg) changes by a specific amount. For this situation, the coefficient for drat is 5.666624. There is also an interaction term between hp and drat, which has a coefficient of -0.033246. So, we can estimate how fuel economy changes for each unit increase in rear axle ratio with a car that has 160 horsepower by the following calculation:

For each unit increase in rear axle ratio, the fuel economy increases by approximately 0.35 mpg for a car with 160 horsepower.

A graph with purple dots

Description automatically generated

The scatter plot of residuals against fitted values shows a random distribution of residuals around the horizontal axis at 0, with no clear pattern or systematic structure. This indicates that the variance of the residuals is constant across all levels of fitted values, supporting the assumption of homoscedasticity. Additionally, there are no obvious outliers or influential points, further validating the model's reliability.

A graph of a normal q-q plot

Description automatically generated

The Normal Q-Q Plot demonstrates that the residuals are approximately normally distributed, as most points closely follow the straight line. While there are slight deviations at the tails, these are not severe enough to significantly violate the normality assumption.

### Evaluating Model Significance

To check if the regression model is significant at a 5% level, we conduct an F-test. The F-statistic is 23.8 with a p-value of 6.098e-09. Since the p-value is much lower than 0.05, we reject the null hypothesis. This means the model is significant, and at least one predictor helps explain the variability in the dependent variable (mpg).

Next, we examine which predictors are significant at a 5% level through t-tests. The null hypothesis states that the predictor’s coefficient is zero, meaning it does not contribute to the model. Conversely, the alternative hypothesis posits that the predictor’s coefficient is not zero, indicating that it does contribute to the model. From the results, the p-value for horsepower (hp) is 0.01175, which is less than 0.05, making it significant. Quarter mile time (qsec) has a p-value of 0.04043, also less than 0.05, so it is significant. Rear axle ratio (drat) has a p-value of 0.03262, marking it significant as well. The interaction term hp:qsec has a p-value of 0.00307, making it significant, while hp:drat, with a p-value of 0.08405, is not significant.

In conclusion, the F-test confirms that the regression model is significant at the 5% level. The individual tests show that hp, qsec, drat, and the interaction term hp:qsec are significant, while hp:drat is not. This helps us understand which predictors are important for explaining variability in mpg.

### Making Predictions Using the Model

For a car with 175 horsepower, 14.2 quarter mile time, and a 3.91 rear-axle ratio, the predicted fuel economy is 21.53 miles per gallon. The 95% prediction interval for the fuel economy of this car ranges from 15.01 to 27.87 miles per gallon, meaning that we can be 95% confident that the actual fuel economy for an individual car with these specifications will fall within this range. Additionally, the 95% confidence interval for the average fuel economy of cars with these specifications ranges from 15.86 to 24.41 miles per gallon, indicating that we can be 95% confident that the true average fuel economy for all cars with these specifications falls within this range.

## 4. Model with Interaction Term and Qualitative Predictor

### Reporting Results

*The general form of the regression model is:*

*Prediction equation from the output can be written as:*

R-squared (R²) value of the regression model is 0.8327 indicates that around 83.27% of the variability in fuel economy (mpg) is explained by the independent variables horsepower(hp), quarter mile time(qsec), their interaction, and the number of cylinders(cyl). This suggests a good model fit. The Adjusted R-squared (R²a) value is 0.8005, which accounts for the number of predictors, indicates that about 80.05% of the variability in mpg is explained, confirming the model's robustness despite its complexity. Overall, these statistics demonstrate the model's effectiveness in relating the predictors to fuel economy.

*A graph with purple dots

Description automatically generated*

In the scatterplot, the residuals are scattered randomly around the horizontal line at Residual = 0, with no discernible pattern or systematic structure. This randomness suggests that the variance of the residuals is constant across all levels of fitted values, indicating that the assumption of homoscedasticity is met.

*A graph with purple dots

Description automatically generated*

In the QQ Normal plot the points in the plot mostly follow the diagonal line, indicating that the residuals are approximately normally distributed. While there are slight deviations at the tails, these deviations are not severe enough to significantly violate the normality assumption.

In conclusion, the assumptions of homoscedasticity and normality of the residuals appear to be satisfied. The Residuals vs. Fitted Values plot indicates no clear pattern, suggesting constant variance, while the Normal Q-Q Plot shows the residuals are roughly normally distributed with minor tail deviations. Overall, these results support the validity of the model's assumptions.

### Evaluating Model Significance

Evaluating the regression model's significance involves two key tests: the overall F-test and individual beta tests. The overall F-test assesses the null hypothesis that the model with all predictors is not significant (i.e., all regression coefficients are zero) against the alternative hypothesis that at least one regression coefficient is not zero. With an F-statistic of 25.88 and a P-value of 2.526e-09, which is significantly less than 0.05, we reject the null hypothesis. This indicates that the overall model is indeed significant at the 5% level of significance.

Moving to the individual beta tests, the null hypothesis for each term is that the coefficient is equal to zero, with the alternative hypothesis being that the coefficient is not zero. The intercept, horsepower (hp), and quarter mile time (qsec) are not significant at the 5% level with P-values of 0.0745, 0.0848, and 0.4828, respectively. However, the term for six cylinders (cyl6) is significant with a P-value of 0.0118, and the interaction term between horsepower and quarter mile time (hp:qsec) is significant with a P-value of 0.0246. Thus, at the 5% level of significance, the significant terms in the model are cyl6 and hp:qsec. This analysis highlights that while the model as a whole is significant, not all individual predictors contribute equally to its explanatory power.

### Making Predictions Using the Model

Using the second regression model, the predicted fuel economy for a car with 175 horsepower, 14.2 quarter mile time, and 6 cylinders is approximately 21.34 miles per us gallon (mpg). The 95% prediction interval for this car's fuel economy ranges from 14.88 to 27.81 MPG. This interval suggests that we can be 95% confident that the actual fuel economy of an individual car with these characteristics will fall within this range. Meanwhile, the 95% confidence interval for the predicted mean fuel economy for cars with these same characteristics is between 18.00 and 24.69 MPG, providing a range where we can be 95% confident that the true mean fuel economy for such cars lies. The prediction interval is wider than the confidence interval because it accounts for the variability in individual observations as well as the uncertainty in estimating the mean response. Therefore, the prediction interval reflects both the inherent variability in individual car performances and the overall model uncertainty, whereas the confidence interval only reflects the uncertainty in estimating the average fuel economy for similar cars.

## 5. Conclusion

Based on the analyses performed, the recommended model is the second regression model, which includes horsepower (hp), quarter mile time (qsec), rear axle ratio (drat), and their interactions. This model is preferred due to its high R-squared value of 0.8327, indicating that approximately 83.27% of the variability in fuel economy (mpg) can be explained by the independent variables, including their interactions and the number of cylinders (cyl). The adjusted R-squared value of 0.8005 further supports the robustness of the model, as it accounts for the number of predictors and provides a more accurate measure of the model's goodness of fit.

The high R-squared and adjusted R-squared values indicate that the model is effective in capturing the relationship between the predictors and fuel economy. In practical terms, this means that the model can reliably predict the mpg based on the specified predictors. The statistical significance of the predictors, as indicated by the p-values from the t-tests, confirms that horsepower, quarter mile time, rear axle ratio, and the interaction between horsepower and quarter mile time significantly influence fuel economy.

The practical importance of these analyses lies in their implications for various stakeholders. For car manufacturers, the insights gained from the model can inform the design and engineering of more fuel-efficient vehicles. Understanding how horsepower, quarter mile time, rear axle ratio, and their interactions impact fuel economy enables manufacturers to make data-driven decisions to improve vehicle performance. For consumers, the model provides valuable information for making informed decisions when purchasing cars, especially those prioritizing fuel efficiency. Policymakers can also benefit from these analyses by using the results to develop regulations and standards aimed at reducing fuel consumption and emissions, contributing to environmental sustainability.

Overall, the analyses performed demonstrate the model's effectiveness in relating engine performance and acceleration characteristics to fuel economy, providing a robust tool for predicting mpg and making informed decisions in the automotive industry