# MAT 303 Module Two Problem Set Report

Interaction Terms and Qualitative Predictors

Yaqub Mohamud

yaqub.mohamud@snhu.edu

Southern New Hampshire University

## 1. Introduction

​​This report​ is an examination of the mtcars data set. The results obtained may be utilized to predict how well a car will perform on the road and will enable us to examine how ​​a car's fuel efficiency​​ (mpg) relates to ​​its horsepower​​ (hp), rear axle ratio (drat), and how this translates into its quarter-mile time (qsec) is an indicator of the car's acceleration capability​​.​ ​Multiple regression with interaction terms and​ the addition of qualitative predictors is the sort of analysis that will be carried out on this data set.

## 2. Data Preparation

In this data collection, the most crucial factors that we will be examining are horsepower, drat, qsec, and mpg. Our objective is to determine if horsepower, qsec, and drat—the independent variables—will have a statistically significant impact on mpg. There are twelve columns, one for each variable pertaining to cars, and 32 rows of data overall, one for each automobile in the dataset.​

## 3. Model with Interaction Term

### Correlation Analysis

*A screenshot of a computer

Description automatically generated*

The numbers above show that there is a -0.7762 association between horsepower and mpg. This indicates that there is a strong ​negative​ association between horsepower ​and ​mpg. Fuel economy tends to decline​ as horsepower rises. ​This implies that​ greater horsepower and power usually translate into higher fuel consumption and worse mpg​. There is a ​0.4187​ correlation between mpg and qsec​. This​ indicates that ​there is a moderate​ positive association ​between quarter-mile time and​ mpg. Fuel efficiency ​tends to improve as quarter-mile time increases​, which indicates slower acceleration. This could suggest that cars that accelerate more slowly​ (higher qsec) might be more fuel-efficient, possibly due to less aggressive engine behavior. Finally, there is a ​0.6812​ association between rear axle ratio and mpg. ​This demonstrates that there​ is a moderate to strong positive association between rear axle ratio and mpg. ​Fuel economy tends to improve as the rear axle ratio rises​​. This​ may indicate that cars with ​higher rear axle ratios (which​ typically have lower engine revolutions at cruising speeds) ​often have higher fuel efficiency​.

### 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)

R² = 0.8207

Interpretation: This indicates that roughly ​82.07% of the​ variation in the dependent variable, ​fuel economy (mpg), can be explained by the model​. It indicates ​a strong relationship​ between the independent variables (such as ​horsepower, quarter mile time, and rear​ ​axle ratio) and the​ fuel economy, suggesting the model fits the data well.

However, the number of predictors in the model is not taken into consideration by R²​​, so adding more variables can artificially increase this value, even if those additional predictors are not particularly useful.

Adjusted R² = 0.7862

Interpretation: The Adjusted R² value ​is a more​ trustworthy indicator ​of model fit​, particularly when working with several predictors. It ​adjusts ​the R² value to account for​​ the number of predictors in the model, preventing overfitting. ​The adjusted value of​ 0.7862 indicates that, following the number of predictors, about ​​78.62% of the​​ variability in fuel economy ​is still​ accounted for by the model.

This means the model still does a great job explaining the data, and the reduction in explanatory power (compared to R²) reflects the fact that the model penalizes complexity, which helps prevent overfitting.

Change in Fuel Economy (mpg) with respect to Quarter Mile Time (qsec) for a car with 160 horsepower (hp): The change in fuel economy due to a one-second increase in quarter mile time is calculated as:

Change in mpg = 1.509555 + (−0.018723 × 160) = 1.509555 − 2.99568 = −1.486125

Interpretation: For each one-second increase in quarter mile time (qsec), the fuel economy decreases by approximately 1.49 mpg for a car with 160 horsepower.

Change in Fuel Economy (mpg) with respect to Rear Axle Ratio (drat) for a car with 160 horsepower (hp): The change in fuel economy due to a one-unit increase in rear axle ratio is calculated as:

Change in mpg = 5.666624 + (−0.033246 × 160) = 5.666624 − 5.31936 = 0.347264

Interpretation: For each one-unit increase in rear axle ratio (drat), the fuel economy increases by approximately 0.35 mpg for a car with 160 horsepower.

A graph with red dots

Description automatically generated

There is no apparent pattern or systematic structure to the scatter plot of residuals against fitted values, which displays a random distribution of residuals around the horizontal axis at 0. This supports the homoscedasticity assumption by showing that the residuals' variance is constant across all levels of fitted values. Furthermore, the model's dependability is further confirmed by the absence of any clear outliers or significant points.

A graph with red dots and blue line

Description automatically generated

The Normal Q-Q Plot shows that most of the points closely resemble the straight line, indicating that the residuals are roughly normally distributed. Although there are some variations at the tails, they are not substantial enough to seriously challenge the normality assumption.

### Evaluating Model Significance

F-statistic = 23.8: This is the ratio of the variation that the regression model explains to the variance that the model is unable to explain. A greater F-statistic indicates that the model fits the data more accurately than the null model, which is a model without any predictors.

​​P-value = 6.098e-09: The​ likelihood of seeing an F-statistic as severe as 23.8 is thus, assuming that the null hypothesis, according to ​which the​ model ​is not significant​, is true. The null hypothesis is rejected as ​the p-value is​ far ​less than the significance​ level of 0.05.

Hypotheses:

Null hypothesis (H₀): The predictor's coefficient is zero, meaning the predictor does not contribute significantly to explaining the dependent variable.

Alternative hypothesis (H₁): The predictor's coefficient is not zero, meaning the predictor does contribute significantly.

Interpretation of p-values:

Horsepower (hp) - p-value = 0.01175: We reject the null hypothesis since​ this p-value ​is less than 0.05 and​ determine that horsepower is a statistically significant predictor of mpg.

Quarter mile time (qsec) - p-value = 0.04043: This p-value is also less than 0.05, so we reject the null hypothesis and conclude that quarter mile time is a significant predictor of mpg.

Rear axle ratio (drat) - p-value = 0.03262: This p-value is well below 0.05, so we reject the null hypothesis and conclude that way horsepower interacts with quarter mile time significantly impacts mpg.

Interaction term (hp:qsec) - p-value = 0.00307: This p-value is well below 0.05, so we reject the null hypothesis and conclude that the interaction between horsepower and quarter mile time significantly impacts mpg.

Interaction term (hp:drat) - p-value = 0.08405: Since this p-value is greater than 0.05, we fail to reject the null hypothesis. This means the interaction between horsepower and rear axle ratio is not statistically significant in explaining mpg.

To sum up, the model is overall significant according to the F-test, and the t-tests reveal which specific predictors (and interactions) are most important for explaining mpg. Based on the results, you can focus on horsepower, quarter mile time, rear axle ​ratio, as well as how horsepower and quarter-mile time interact​ as the key factors affecting mpg. The interaction between horsepower and rear axle ratio, however, does not provide significant explanatory power.

### Making Predictions Using the Model

Predicted Fuel Economy:

With a quarter-mile timing of 14.2 seconds, 175 horsepower, and a rear-axle ratio of 3.91, the car should get 21.53 miles per gallon (MPG).

95% Prediction Interval:

This interval (15.01 to 27.87 MPG) explains the variations among specific cars. It shows that we have a 95% confidence level that any given cars with these specs will possess fuel economy that is in line with this range.

95% Confidence Interval for the Mean:

This narrower range (15.86 to 24.41 MPG) represents the average fuel economy of all ​cars that meet these requirements. It​ means we are 95% confident ​that this range represents the average MPG for this class of cars.​

## 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:

The R-squared (R²) value of 0.8327 for the regression model indicates that the variables horsepower (hp), quarter-mile time (qsec), their interaction, and the number of cylinders (cyl) account for around 83.27% of the variation in fuel efficiency (mpg). This suggests that the response variable and the predictors are well matched.

The Adjusted R-squared (R²a) ​value is 0​.8005, account for the number of predictors, reducing ​overfitting ​possibility. With ​80.05% of the variability​ still explained, this confirms the model's robustness despite its complexity.

A diagram with red dots

Description automatically generated

There is no obvious pattern or systematic structure to the scatterplot; instead, the residuals are arbitrary with respect to the horizontal line at Residual = 0. The requirement of homoscedasticity is satisfied as this randomness implies that the variance of the residuals is constant across all levels of fitted values.

A graph with red dots

Description automatically generated

The residuals are roughly normally distributed, as indicated by the QQ Normal plot's dots, which primarily ​observe the diagonal line​. Even if there are some variations at the tails, they are not substantial enough to seriously breach the normality assumption.

In summary, the homoscedasticity and normality requirements for the residuals appear to be satisfied. Despite the lack of a discernible pattern in the Residuals vs. Fitted Values plot, the Normal Q-Q Plot indicates that the residuals are continuously variable and nearly normally distributed with very little tail deviations. These results, taken together, give the model's assumptions validity.

### Evaluating Model Significance

We use both the overall F-test and individual beta tests to determine whether the regression model is significant at the 5% significance level. The total F-test result indicates that a significant portion of the variation in fuel economy (mpg) may be explained by the predictors (horsepower, quarter-mile time, interaction term, and cylinder indicators) combined. The null hypothesis is rejected since the F-statistic is 25.88 and the P-value is 2.526e-09, both of which are substantially below 0.05. At the 5% level of significance, this suggests that the model is typically statistically significant.

The alternative ​hypothesis for​ the individual beta tests is that the ​coefficient is​ not zero, whereas the null ​hypothesis is that the coefficient​ for every term is equal to zero.

Significant Predictors:

Cyl6 (6 cylinders): With a P-value of 0.0118 (< 0.05), this term is significant, indicating that the presence of 6 ​cylinders significantly affects fuel efficiency​.

Interaction Term (hp:qsec): With a P-value of 0.0246 (< 0.05), this term is also significant, highlighting that the result of combining ​horsepower with quarter-mile time​ is important in explaining variability in mpg.

Non-significant Predictors:

​​At the 5% level, the intercept, horsepower (hp), and quarter-mile time (qsec​) do not each substantially contribute since ​their P-values are​ more than 0.05.

In summary, whereas the model as a whole is important (from the F-test), not all predictors are equally important. ​​The significant terms (cyl6 and hp:qsec​​) suggest specific factors that strongly influence fuel economy, while others (e.g., hp and qsec individually) may have weaker or indirect effects.

**Making Predictions Using the Model**

Predicted Fuel Economy:

For a car with six cylinders, 175 horsepower, and a 14.2 quarter-mile time, the estimated fuel economy is approximately 21.34 MPG.

95% Prediction Interval: 14.88 to 27.81 MPG.

Interpretation:

This interval captures the range in ​which we​ anticipate ​the fuel efficiency of​​​ an individual ​car with these​ characteristics ​​to fall, with 95% confidence​​.

95% Confidence Interval for the Mean: 18.0018.0018.00 to 24.6924.6924.69 MPG.

Interpretation:

This ​interval shows the location of the actual mean fuel​ economy for all cars with these ​characteristics is​ expected to lie, ​with 95% confidence.​

The prediction ​interval is​ greater ​than the confidence interval​ as the PI takes into consideration the unpredictability of individual data (random error), while The CI ​solely takes into consideration uncertainty when calculating the average answer​. This difference emphasizes the additional uncertainty when predicting outcomes for individual cars versus estimating ​the average fuel efficiency for​ all cars with the same details.

## 5. Conclusion

Considering ​the results of the analyses​, the second ​regression model is recommended​ for predicting fuel economy (mpg) ​due to its​ robust statistical performance and practical relevance. This comprises the ​rear axle ratio (drat), horsepower (hp), quarter-​​mile time (qsec), and​ their interrelationships. This model ​shows that roughly​ 83.27% of ​model's predictions explain ​the variation in fuel efficiency​​​. The model's robustness and good generalization are confirmed by ​the adjusted R-squared value of 0.8005​, which explains ​the number of predictors.​

The model is successful in demonstrating how the predictions and fuel economy are related, as evidenced by the high R-squared and modified ​​R-squared values​​. Practically speaking, this indicates that the model can accurately forecast the mpg using the given predictors. ​The p-values from​ the t-tests demonstrate the statistical significance of the predictors, confirming that ​fuel efficiency is​ strongly influenced by ​​horsepower, rear axle ratio, quarter mile time​, and how these factors interact.​

The effects of these studies for different stakeholders are what give them practical significance. Car makers may use the model's observations ​to ​guide the​​ engineering and ​design ​of more fuel-efficient​​ automobiles. Manufacturers may ​make judgments based on data to​ enhance ​vehicle efficiency​ by knowing how horsepower, quarter mile ​​time, rear axle ratio, and how they​ interact​ affect fuel efficiency. The model gives buyers useful information to help them make well-informed decisions when buying automobiles, particularly for those who prioritize fuel efficiency. These assessments can also be useful to policymakers, who can use the findings to create rules and guidelines that will lower emissions and fuel use and promote environmental sustainability.

In conclusion, the second regression model demonstrates strong explanatory power, as reflected in its high R-squared and ​​Adjusted R-squared, and​​ statistical significance of its predictors. Its practical relevance for manufacturers, consumers, and policymakers underscores its value as a tool for predicting fuel economy and directing choices in the automobile sector.​​