# MAT 303 Module One Problem Set Report

Multiple Regression

Andrea Plunkett Jackson

[andrea.jackson11@snhu.edu](mailto:andrea.jackson11@snhu.edu)

Southern New Hampshire University

## Introduction

In this study, I am delving into the mtcars dataset, a comprehensive collection of automobile-related variables. This dataset encompasses variables such as miles per gallon (mpg), number of cylinders (cyl), displacement (disp), horsepower (hp), rear axle ratio (drat), weight (wt), quarter mile time (qsec), engine configuration (vs), transmission type (am), and number of gears (gear). With the focus on how a car’s rear axle ratio and horsepower relate to its fuel efficiency.

The analysis of this dataset can offer valuable insights, particularly in understanding the factors that contribute to fuel economy. By creating predictive models, it is possible to identify key features that enhance fuel efficiency. These features can then be incorporated into vehicle designs, potentially leading to the production of vehicles with superior fuel efficiency. This would be beneficial for both the environment and the consumer.

The analysis of this problem set will encompass several types of statistical methods. Initially, a correlation analysis will be performed to discern the associations between different variables, specifically using Pearson correlation coefficients to quantify the degree of linear relationship between pairs of variables.

Subsequently, we will employ a multiple regression model, which is a type of predictive modeling. This model will be used to predict the fuel economy based on the other variables in the data set. The selection of the predictive model will hinge on the nature of the relationships identified during the analysis.

## Data Preparation

The key variables I will be examining are rear axle ratio, horsepower, and miles per gallon. The full dataset consists of 11 columns, each representing different variables associated with the vehicles. There are 32 rows in total, each corresponding to a unique vehicle included in the dataset. The focus of my analysis will be on the relationships between fuel efficiency and rear axle ratio, and between fuel efficiency and horsepower. This will provide insights into how these variables interact and their potential impact on a vehicle’s fuel efficiency.

## Multiple Regression Model

### Correlation Analysis

The scatterplot and correlation matrix was used to visualize and provide valuable insights into the relationships between Fuel Efficiency, Rear Axle Ratio, and Horsepower.

• Fuel Efficiency and Rear Axle Ratio: The scatterplot suggests a moderate, positive relationship between these two variables. As Fuel Efficiency increases, the Rear Axle Ratio also seems to increase. However, the relationship does not appear to be strong or linear, indicating that other factors also influence a vehicle’s fuel efficiency. The Pearson Correlation Coefficient between these two variables is approximately +0.681, confirming the moderate positive correlation. This suggests that while there may be some correlation between a vehicle’s rear axle ratio and its fuel efficiency, other factors also play a significant role in determining a vehicle’s fuel efficiency.

A graph with red dots

Description automatically generated

• Fuel Efficiency and Horsepower: The scatterplot illustrates a negative relationship between these two variables. As Fuel Efficiency increases, Horsepower seems to decrease, and vice versa. The Pearson Correlation Coefficient between these two variables is approximately -0.776, indicating a moderate negative correlation. This suggests that vehicles with higher fuel efficiency may have less horsepower, and those with more horsepower may be less fuel-efficient. However, the points are quite dispersed, indicating that other factors may also influence these variables.

A graph with red dots

Description automatically generated

These coefficients suggest that while there are observable relationships between Fuel Efficiency, Rear Axle Ratio, and Horsepower, these relationships are moderate and influenced by other factors. Further statistical analysis would be needed to confirm these observations and determine the exact nature of the relationships between these variables.

### Reporting Results

The general form of a multiple regression model is:

For this specific case where fuel efficiency (mpg) is the response variable and rear axle ratio (drat) and horsepower (hp) are predictors, the prediction equation from the provided R script output would be:

This equation allows us to predict the fuel efficiency of a vehicle given its rear axle ratio and horsepower. For example, we can choose a vehicle from our dataset and input that vehicle’s rear axle ratio into our equation for () and horsepower for (Doing this would allow us to calculate the predicted fuel efficiency of the chosen vehicle.

The () (R-squared) value is 0.7412 and the (Adjusted R-squared) value is 0.7233. This tells us that approximately 74% of the variation in fuel economy can be explained by a model that uses predictors for rear axle ratio and horsepower. The adjusted R-squared value considers the number of predictors in the model and is a more accurate measure of the goodness of fit.

The beta estimates are ({drat} = 4.059187) for rear axle ratio and ({hp} = -0.051787) for horsepower. This means that for each 1 unit increase in rear axle ratio, our predicted fuel efficiency will increase by 4.059187 units on average, holding horsepower constant. Conversely, for each 1 unit increase in horsepower, our predicted fuel efficiency will decrease by 0.051787 units on average, holding rear axle ratio constant. These estimates provide valuable insights into how changes in rear axle ratio and horsepower might affect a vehicle’s fuel efficiency.

The scatterplot below of residuals against fitted values provides a visual representation of these residuals and their relationship with the fitted values. This can provide valuable insights into the accuracy of the model’s predictions and guide further refinements to improve its predictive accuracy. It also highlights the potential complexities in predicting fuel efficiency, which may be influenced by a range of factors and may not follow a simple linear relationship with predictors like rear axle ratio and horsepower.

A graph with red dots

Description automatically generated

A graph of a normal q-q plot

Description automatically generated

In the scatter plot titled “Residuals against Fitted Values”, the residuals are randomly scattered around zero across the range of fitted values, it suggests that the variance of the residuals is constant, hence, homoscedasticity holds.

Based on the Normal Q-Q Plot the residuals seem to deviate from the reference line at both ends of the plot, suggesting potential deviations from normality, especially in the tails. This could indicate potential outliers or heavy tailenders in the distribution of residuals. However, the majority of points do follow the trend of the reference line, indicating some adherence to normality.

### Evaluating Model Significance

For evaluating model significance of a multiple regression model and its predictors, specifically rear axle ratio and horsepower, in predicting fuel efficiency. I am using a 5% significance level for these tests. For the overall model, the equation for our overall F-test is:

I am checking if at least one predictor variable significantly affects fuel efficiency. The P-value obtained is extremely small (9.109e-12), which is much less than the 5% significance level. This leads me to reject the null hypothesis (that all predictor variables have no effect) and accept the alternative hypothesis (that at least one predictor variable does have an effect). In simpler terms, this means that either rear axle ratio, horsepower, or both have a significant impact on fuel efficiency.

Next, I individually test each predictor variable for significance. I find that both rear axle ratio and horsepower have confidence intervals that do not include zero, indicating they are significant contributors to the model at a 5% level. The equation for this is:

=0

Specifically, the confidence interval for horsepower ranges from -0.0708 to -0.0328, and for rear axle ratio, it ranges from 2.2610 to 7.1353. This suggests that both rear axle ratio and horsepower have a significant relationship with fuel efficiency.

The 95% confidence intervals for the parameter estimates of rear axle ratio and horsepower are as follows:

For horsepower (hp), the 95% confidence interval is from -0.0708 to -0.0328. This means we can be 95% confident that the true value of the effect of horsepower on fuel efficiency lies within this range. Since both ends of the interval are negative, it suggests that as horsepower increases, fuel efficiency decreases.

For rear axle ratio (drat), the 95% confidence interval is from 2.2610 to 7.1353. This means we can be 95% confident that the true value of the effect of rear axle ratio on fuel efficiency lies within this range. Since both ends of the interval are positive, it suggests that as rear axle ratio increases, fuel efficiency also increases.

### Making Predictions Using the Model

The given model for predicting fuel efficiency (mpg) is:

mpg=10.78961+4.059187−0.051787

where:

* is the rear axle ratio (drat)
* is the horsepower (hp)

Let us use this model to predict the fuel efficiency for a car with a rear axle ratio of 3.15 ( = 3.15) and a horsepower of 120 ( = 120).

Substituting these values into the model, we get:

mpg=10.78961+4.059187∗3.15−0.051787∗120

mpg = 23.01729

Now, let us calculate the residual. The residual is the difference between the observed value and the predicted value. In this case, the observed value is 20.5 miles per gallon.

So, the residual would be:

residual=observed mpg−predicted mpg

residual = 23.01729 -20.5

So, our residual is 2.517288. This residual represents the difference between the observed fuel efficiency and the fuel efficiency predicted by our model. A positive residual indicates that the actual value is higher than the predicted value. In this case, the positive residual of 2.517288 suggests that the actual fuel efficiency of the car is higher than what our model predicted.

To estimate the fuel efficiency of a car with a rear axle ratio of 3.15 and a horsepower of 120, we use a statistical tool known as the 95% prediction interval. In this case, the interval is roughly (12.6449, 26.1045). This suggests that, based on our existing data, there is a 95% likelihood that the car’s fuel efficiency will lie within this range.

In contrast, the 95% confidence interval, approximately (17.5736, 21.1777), gives us a range in which we can be 95% sure that the true average fuel efficiency of a car (with a rear axle ratio of 3.15 and horsepower of 120) lies.

The prediction interval is typically wider than the confidence interval due to the different types of uncertainty each one considers. The confidence interval (17.5716, 21.1777) in this case, reflects the uncertainty associated with estimating the average fuel economy of cars with a rear axle ratio of 3.15 and horsepower of 120. It provides a range in which we can be 95% confident that the true average fuel economy lies.

On the other hand, the prediction interval (12.6449, 26.1045) not only considers this uncertainty, but also the variability among individual cars. This means it accounts for the fact that even cars with the same rear axle ratio and horsepower can have different fuel economies due to other factors. Therefore, the prediction interval provides a range that is expected to contain the fuel economy of a new car of the same specifications, making it wider than the confidence interval.

## Conclusion

In conclusion, based on the conducted analysis and assuming a large enough sample size, I would recommend this model. It accurately demonstrates that the rear axle ratio and horsepower have a significant impact on fuel efficiency. The Pearson Correlation Coefficient between these variables is approximately +0.681, indicating a moderate positive correlation. This suggests that while the rear axle ratio of a vehicle may influence its fuel efficiency, other factors also play a significant role.

The R-squared (R^2) value is 0.7412, and the Adjusted R-squared (R^2a) value is 0.7233. This indicates that about 74% of the variation in fuel economy can be explained by a model that uses rear axle ratio and horsepower as predictors. The adjusted R-squared value, which considers the number of predictors in the model, provides a more accurate measure of the model’s fit.

The beta estimates are (b\_1{drat} = 4.059187) for rear axle ratio and (b\_2{hp} = -0.051787) for horsepower. This means that for each 1 unit increase in rear axle ratio, our predicted fuel efficiency will increase by 4.059187 units on average, holding horsepower constant. Conversely, for each 1 unit increase in horsepower, our predicted fuel efficiency will decrease by 0.051787 units on average, holding rear axle ratio constant.

Our residual is 2.517288. This residual represents the difference between the observed fuel efficiency and the fuel efficiency predicted by our model. A positive residual indicates that the actual value is higher than the predicted value. In this case, the positive residual of 2.517288 suggests that the actual fuel efficiency of the car is higher than what our model predicted.

The practical significance of this analysis is that it assists car manufacturers in predicting the fuel efficiency of vehicles in the design phase. They can use the rear axle ratio of the vehicle and the horsepower it will have to calculate how fuel efficient it will be. This will be useful to know so they do not waste capital investment on making a car that will not be appealing to their customers. By leveraging predictive information during the design phase, automakers can proactively adjust. This way, when they reach the production phase, modifications have already been implemented. As a result, they can optimize resources, reduce costs, and focus on producing appealing cars.

## Citations

MAT 303 Module 1. (n.d.). Www.youtube.com. https://www.youtube.com/watch?v=tZA3SAvp\_Lc

‌Ahmad Qadri. (2020, March 7). MAT 303 Module 1 Problem Set. YouTube. https://www.youtube.com/watch?v=7PQ-ROTfF1U

‌mtcars function - RDocumentation. (n.d.). Www.rdocumentation.org. https://www.rdocumentation.org/packages/datasets/versions/3.6.2/topics/mtcars

‌