# MAT 303 Module One Problem Set Report

Multiple Regression

Ashley De Venuto

ashley.devenuto@snhu.edu

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## Introduction

The dataset that is being explored for this analysis is based on fuel efficiency. The data being analyzed is rear axel ration and horsepower. The focus of the dataset is to see if horsepower and/or rear axel ration effects the fuel efficiency and miles per gallon the vehicle can achieve. The results of the analysis will help the auto manufacturer determine what the right rear axel ration and horsepower should be to maximize a vehicles mile per gallon. The results can also help by driving more customers to a certain type of vehicle because of its fuel efficiency which in turn will increase sales. With this analysis, a multiple regression model is used. Scatterplots are used to show associations between fuel efficiency and rear axel ration as well as fuel efficiency and horsepower. There will also be analysis on the correlation coefficients between the dependent variable (fuel economy) and the independent variables (rear axel ration and horsepower).

## Data Preparation

There are many different variables within this dataset. For this analysis, the focus is on fuel economy (mpg) as the dependent variable and rear axel ratio (drat) and horsepower (hp) as the independent variables. There are 12 columns and 32 rows within this dataset. These columns represent the variables, and the rows represent the type of car.

## Multiple Regression Model

### Correlation Analysis

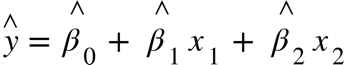
*A graph with red dots

Description automatically generated* A graph with red dots

Description automatically generated

When using scatterplots, it’s easier to see if there’s any correlation between variables, here are the scatterplots for both the mpg against drat and mpg against horsepower. For the scatterplot for the mpg against drat, it looks like when the rear axle ratio increases so does the fuel economy. When the scatterplot for mpg against hp is examined, it looks like when the horsepower increases, miles per hour decreases. It makes sense that mpg against horsepower has a negative correlation because horsepower burns fuel quickly. Horsepower has a -0.7762 correlation with fuel economy. Rear axle ration has a positive correlation with fuel economy. Rear axle ration has a 0.6812 correlation with fuel economy.

### Reporting Results

The general form of the multiple regression model is {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mi>E</mi><mfenced><mi>Y</mi></mfenced><mo>&#xA0;</mo><mo>=</mo><mo>&#xA0;</mo><msub><mi>&#x3B2;</mi><mn>0</mn></msub><mo>&#xA0;</mo><mo>+</mo><mo>&#xA0;</mo><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><mo>+</mo><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub></mstyle></math>","origin":"MathType for Microsoft Add-in"}. The Y indicates the fuel economy, x1 represents the rear axel ratio, and x2 represents horsepower. The prediction regression model is . The multiple regression model equation when filling in with fuel efficiency, rear axle ration, and horsepower is {"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>&#xA0;</mo><mo>=</mo><mo>&#xA0;</mo><mn>10</mn><mo>.</mo><mn>789861</mn><mo>&#xA0;</mo><mo>+</mo><mo>&#xA0;</mo><mn>4</mn><mo>.</mo><mn>698158</mn><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><mo>-</mo><mo>&#xA0;</mo><mn>0</mn><mo>.</mo><mn>051787</mn><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub></mstyle></math>","origin":"MathType for Microsoft Add-in"}. The value for and the value of . The value for indicates that there is a 74.12 % of the fuel economy can be explained by rear axle ration and horsepower. This also indicates a strong linear link a vehicle’s fuel economy and their rear axle ration and horsepower. This link indicates that rear axle ration and horsepower are reliable variables to determine fuel economy. When it comes to , there’s not much a difference. If other models were made with these same variables, there would be a very small impact with results. According to the coefficient of the rear axle ration, the vehicle’s fuel economy will increase by 4.698158 rotations for every one rotation of the wheel axle. Horsepower will decrease fuel efficiency by 0.051787 due to the negative Infront of the number. A fitted value is based on the value of the dependent variable. The residual value is determined by the discrepancy between the actual value and anticipated value of the dependent variable. When trying to indicate the assumptions of homoscedasticity, a scatterplot was used where residuals were against fitted values. Looking at the scatterplot below indicates the residuals for both rear axle ratio and horsepower predictor variables have a constant variance. This in turn also means that the points on the scatterplot have near equal distance s from the horizontal line at zero. The Q-Q plot analyzes the residuals normalcy. The Q-Q plot below indicates that that the residuals are normally distributed. There are four sites that significantly deviate from the diagonal line. Besides the four sites that deviate from line, the results show that the residuals are normal.

A diagram with red dots

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

Description automatically generated

### Evaluating Model Significance

An F-test was run to determine if the model is significant at 5% level of significance. First, the null hypothesis, alternative hypothesis and P-value needed to be identified. The F-test was used to determine if a linear relationship existed between at least one of the predictor variables. The null hypothesis equation is and the alternative hypothesis equation would be . The P-value is 3.081e-09. Due to the P-value being close to zero, but not equal to zero, I rejected the null hypothesis in favor of the alternative hypothesis. This means that at least one of the predictor variables is linearly related to Y (dependent variable). A t-test would be carried out on each of the predictor variables to determine which one is significant at 5%. The null hypothesis for the t-test is and the alternative hypothesis equation is . The first variable looked at is rear axle ratio (drat). Drat t value = 3.943 and the p-value = 0.000467. I would reject the null hypothesis in favor of the alternative hypothesis since 0.000467 is less than 0.05. I did the same procedure with horsepower. Horsepower has a t-value = -5.573 and a p-value of 5.17e-06. I would also reject the null hypothesis in favor of the alternative hypothesis because the p-value is less than the 5% significance level. The 95% confidence interval shows where the upper and lower bounds of the interval within each parameter.

### Making Predictions Using the Model

For a vehicle that has a rear axle ratio of 3.15 and has a horsepower of 120, the fuel economy would be 19.3746187. If the car averaged 20.5 miles per gallon, the residual for the observation would be 1.125. The 95% prediction interval is (12.6449 – 26.1045) which means that there is a 95% confidence level within the prediction interval for fuel economy. The confidence interval is (17.5716 – 21.1777) which means that there is a 95% confidence level within the confidence interval. There is a wider prediction interval compared to the confidence interval. This indicates that the predicted variables value along with the response varies due to inaccuracies in the regression.

## Conclusion

After finishing the analysis and seeing the results, I would recommend using this model depending on if the sample size is sufficiently large. When it comes to a car’s rear axle ratio and horsepower, the results make an accurate prediction of the miles per gallon it can achieve. The results show that a car with a greater rear axle ration and a lower horsepower have greater miles per gallon (mpg). Both the rear axle ratio and horsepower have an impact on the fuel economy. A multiple regression model was used. The rear axle ratio and horsepower representing the independent/predictor variables and the fuel economy as the dependent/response variable. I also used scatterplots to show the relationship between fuel economy and rear axle ratio and fuel economy and horsepower. These scatterplots were the visual representation of the relation between all three variables. The results showed that the more rear axle ratio the car had, the more fuel economy (mpg) the car had. The results also showed that the more horsepower the car had, the lee fuel economy (mpg) the car had. Individual t-test and an F-test were also carried out in this analysis. After conducting these tests, it was found that both the predictor variables (drat and hp) were significant and had a significant relationship with the response variable. The importance of the analysis was to give automobile makers evidence on how the rear axle ratio and horsepower had a significant impact on the fuel economy on a car. This evidence would help them make informed decisions on how to design and build cars that appeal to buyers.