# MAT 303 Module One Problem Set Report

Multiple Regression

Sarah Steinbaum

Sarah.steinbaum@snhu.edu

Southern New Hampshire University

## Introduction

The following statistical analyses will be used to determine how a vehicle’s horsepower and rear axle ratio impact the vehicle’s fuel efficiency. Rear axle ratio refers to the number of turns of the drive shaft for every one rotation of the wheel axle. A high rear axle ratio is associated with higher torque. The data set that will be utilized in these analyses is a set of 32 vehicles and 12 variables including the car type, horsepower, and miles per gallon of each vehicle. The results of these analyses may assist the car company to calculate the balance between horsepower and rear axle ratio needed to improve fuel efficiency. A multiple regression model will be utilized to determine the relationship between horsepower and fuel economy as well as rear axle ratio and fuel economy.

## Data Preparation

This data set is compromised of important variables including ‘hp’ for gross horsepower, ‘mpg’ for fuel economy in US miles per gallon, and ‘drat’ for rear axle ratio. There are 12 columns in the data set in which one variable is represented per column. There are 32 rows that each represent a vehicle.

## Multiple Regression Model

### Correlation Analysis

Chart, scatter chart

Description automatically generated

Upon viewing the scatterplot show above, we can see that there is a positive trend associated with fuel efficiency and rear axle ratio. The lower the rear axle ratio, the lower the fuel economy and a vehicle with a higher rear axle ratio has a higher fuel economy. This also means that a vehicle with a lower rear axle ratio is not as fuel efficient compared to a vehicle with a higher rear axle ratio. This considered a positive correlation because as one variable increases, so does the other. The correlation coefficient is strong when its absolute value is between .08 and 1, moderate between 0.4 and 0.8, and considered weak if it is less than 0.4 to zero. The Pearson Correlation Coefficient between fuel efficiency and rear axle ratio is 0.6812 which indicates a moderate positive correlation.

*Chart, scatter chart

Description automatically generated*

The scatterplot above provides a visual of the correlation between horsepower and fuel efficiency. A vehicle with lower horsepower is associated with higher fuel efficiency and a vehicle with higher horsepower is associated with less fuel efficiency. Essentially, a vehicle would require more fuel to power a higher horsepower. This is considered a negative correlation because as one variable increases the other decreases. The Pearson Correlation Coefficient between fuel efficiency and horsepower is -0.7762, the absolute value is 0.7762 which indicates a moderate negative correlation.

### Reporting Results

The general form of the multiple regression model is:

The prediction equation is:

In the prediction equation of multiple regression model, the variable fuel efficiency (mpg) is Y and is the response variable. Rear axle ratio (drat) is and horsepower is , both of which are predictor variables. is considered the slope coefficient for the predictor variables. When the values from the R script are entered, the equation is:

R-squared is the coefficient of multiple determination. The value of R-squared in this model is 0.74112 and is the ability of the model to predict the variation and the response variable. This means that when using the predictors rear axle ratio and horsepower we can explain 74.112% of the variation in fuel economy. The Adjusted R-squared is the adjusted coefficient of multiple determination and its value in this model is 0.7233. The Adjusted R-squared is an adjustment to R-squared that allows alternative models for the same response variable to be compared (Berrier, 2016).

The beta estimates () represent the estimates of the slope coefficients for the predictor variables. They help determine what their affect is of predicting the response. The beta estimate for rear axle ratio in this model is 4.698158. The beta estimate for horsepower is -0.051787. The beta estimates help determine what affect the predictors have on the response variable which in this model is fuel efficiency. Essentially, as the predictor variables increase, the response variable will decrease by the beta value. For example, if the rear axle ratio were to increase by 1 unit, the response variable (fuel efficiency) will decrease by 4.698158.

A fitted value is a predictive value that is the model’s prediction of the mean response. This occurs once all predictor values have been inputted into the model. A residual is the difference between the actual value of the response value and the predicted value.

*Chart, scatter chart

Description automatically generated*

*Chart, scatter chart

Description automatically generated*

The first plot above is the Residual Against Fitted Values Plot which can assist us in making assumptions regarding homoscedasticity. This graph displays the residual values for the predictor variables rear axle ratio and horsepower and seems to portray that the variables have a constant variance. The Normal Q-Q Plot above allows us to determine that the residuals are normally distributed as most of the data points are on or near the linear trendline. If the data points were significantly deviated from the trendline, that would indicate little or no residual normalcy.

### Evaluating Model Significance

An overall F-test can be conducted in order to determine if there is a linear relationship between at least one predictor variable. The null hypothesis is that there is no linear relationship between any of the predictor variables. The alternative hypothesis is that a linear relationship exists between at least one of the predictor variables.

Null Hypothesis

Alternative Hypothesis for *i* = 1, 2

The P-value for this model is 3.081e-09. Since the level of significance is 5% or 0.05 is greater than the P-value, we have sufficient evidence to conclude that the model is statistically significant and can reject the null hypothesis. Now that we know at least one of the predictor variables is not 0, we need to find out which one or both is statistically significant. In order to do this, an individual t-test will need to be conducted for both variables. The null hypothesis for the predictor variables rear axle ratio and horsepower individually is that no linear relationship exists. The alternative hypothesis is that a linear relationship does exists.

Null Hypothesis for *i* = 1, 2

Alternative Hypothesis for *i* = 1, 2

The P-value for rear axle ratio is 0.000467 and the P-value for horsepower is 5.17e-06. Both variables are less than the 5% level of significance which indicates that both are statistically significant. There is sufficient evidence to reject the null hypothesis and conclude that the variables rear axle ratio and horsepower have a significant linear relationship with the response variable, fuel economy. A 95% confidence interval for the parameter estimates of rear axle ratio and horsepower indicates the probability that each parameter falls within the upper and lower bounds of the interval.

### Making Predictions Using the Model

In order to find the predicted fuel efficiency for a car that has a rear axle ratio of 3.15 and a horsepower of 120 we need to use the prediction equation from earlier and place 3.15 into and 120 into :

We find that the predicted fuel efficiency is 19.3747. If a car achieves an average of 20.5 miles per gallon, the residual for this observation would be -1.1253 as we subtract the predicted value from the actual value. The 95% prediction interval for the car in this scenario would be (12.6449 – 26.1045). The 95% confidence interval would also be (17.5716 – 21.1777). The prediction interval is wider than the confidence interval because there is a smaller sample compared to the confidence interval which has a larger sample.

## Conclusion

Based on the analyses performed and a sufficiently large sample size, I would recommend using this model. The model makes an accurate prediction on how many miles per gallon (mph) a vehicle will use based on its rear axle ratio and horsepower. A multiple regression model was developed with rear axle ratio (drat) and horsepower (hp) as the predictor and independent variables and fuel efficiency (mpg) as the response and dependent variable. Scatterplots were created to portray the relationship between rear axle ratio and fuel efficiency as well as horsepower and fuel efficiency which provided a visual to see the relationship between the variables. The multiple regression model and the scatterplots allowed us to determine that a vehicle with a higher rear axle ratio is associated with higher fuel economy and a vehicle with higher horsepower is associated with lower fuel economy. An overall F-test was conducted and determined that at least one of the predictor variables between rear axle ratio and horsepower are statistically significant. After the induvial t-test was conducted, we determined both of the predictor variables are statistically significant. Predictions were then made to find the fuel efficiency based on the values of the predictor variables. The practical importance of the analyses is to demonstrate to vehicle manufacturers the impact rear axle ratio and horsepower have on a vehicle’s fuel efficiency. This can assist the manufacturers to design and build cars according to the fuel efficiency needs of the customers and improve sales.

## Citations

Berrier, J. (2016). MAT 303: Applied Statistics 2 for Science. Zyante Inc. (zyBooks.com)