# MAT 303 Module One Problem Set Report

Multiple Regression

Kayla Sacks

kayla.sacks@snhu.edu

Southern New Hampshire University

## Introduction

*This statistical analysis I will be performing is how the rear axle ratio and horsepower affects a car’s fuel efficiency. The data set that I will be using focuses on how the fuel efficiency, or miles per gallon (mpg), is affected by the rear axle ratio (drat) and horsepower (hp). The results from this analysis can be used to create cars that are more fuel-efficient. By being able to make more fuel-efficient the company will be able to make cars that consumers will be more interested in, which will help to increase sales. For the analysis I will be building a multiple regression model and doing a correlation analysis.*

## Data Preparation

*The important variables in the data set that I will be using are the fuel economy in miles (mpg) for the dependent variable and the rear axle ratio (drat) and horsepower (hp) for the independent variables. There are 32 rows and 12 columns included in the data set.*

## Multiple Regression Model

### Correlation Analysis

*To visualize the relationship between our important variables we will use scatterplots. The first scatterplot below shows the relation between fuel efficiency (miles per gallon) against the rear axle ratio. Although not strong, It shows a positive correlation. We can see that the lower rear axle ratio has lower fuel efficiency. The second scatterplot illustrates the relation between fuel efficiency (miles per gallon) and horsepower and shows a negative correlation. From the plot we can see that the higher the horsepower the lower the fuel efficiency.*

*A diagram of a graph

Description automatically generatedA graph with red dots

Description automatically generated*

*The correlation coefficient between fuel efficiency and rear axle ratio is 0.6812 which indicates that there is a moderate positive correlation. The correlation coefficient between fuel efficiency and horsepower is -0.7762 which shows a strong negative correlation.*

### Reporting Results

*The multiple regression model equation for fuel efficiency (mpg) as the response variable and rear axle ratio (drat) and horsepower (hp) as the predictors is:*

*Based on the output of the multiple regression model, the prediction model equation is below where the the rear axle ratio input is and horsepower is :*

*The (R-squared) value is 0.7412 and the (Adjusted R-squared) value is 0.7233 for the model. This shows that that the regression model explains 74% of the variation in fuel economy using the rear axle ratio and horsepower as predictors. The bata estimates for rear axle ratio and horsepower are 4.6982 and -0.05187 respectively. This means that for each time the predictor variables increase by one, the response variable will increase by 4.6982 and decrease by -0.05187.*

*The fitted value is the predicted value of the dependent variable, or the variable with the residual being the true values of the outcome variable minus the fitted value. After obtained the fitted values and residuals for the dataset and created the scatterplot below for Residuals against Fitted Values.*

*A diagram with red dots

Description automatically generated*

*Below is the Normal Q-Q plot that was created from the residuals.*

*A graph of a normal q-q plot

Description automatically generated*

*Based on the Residual against Fitted Values plot not show any patterns, the model displays homoscedasticity. Due to the points deviating a significant amount on the q-q plot, the plot has non-normally distributed residuals.*

### Evaluating Model Significance

*To evaluate the model significance for the regression model, we will see if the model significance is at a 5% level of significance. To do this I will carry out the overall F-test starting with the null hypothesis and the alternative hypothesis shown below:*

*The P-value is 3.081e-9 which is much lower than our 0.05 level of significance which tells us that there is a statistically significant difference, and we can reject the null hypothesis. Since we reject the null hypothesis, we will use the alternative hypothesis which means that there is a significant linear relationship between fuel efficiency and one or both response variables.*

*To evaluate the variables individually if they are at a 5% significance level, we will do an individual t-test for both rear axle ratio and horsepower. The equations for this test are below with :*

*The p-value for the rear axle ratio (drat) is 0.000467 and the p-value for horsepower (hp) is 5.17e-06. Both values are lower than the 0.05 level of significance so the null hypothesis will be rejected, and the alternate hypothesis will be used. The conclusion of this test tells us that both the rear axle ratio and horsepower have a statistically significant relationship to fuel efficiency. The 95% confidence interval for the parameter estimates is shown below.*

A screenshot of a graph

Description automatically generated

### Making Predictions Using the Model

*To find the predicted fuel efficiency for a car with a rear axle ratio of 3.15 and a horsepower of 120, we will input these values into the prediction model equation. The calculations for the predicted fuel efficiency are:*

*If the car above achieves an average of 20.5 miles per gallon is the Y value, the residual, , for this observation is:*

*The 95% prediction interval and 95% confidence intervals are shown below for a car with a rear axle ratio of 3.15 and horsepower of 120.*

*A screenshot of a graph

Description automatically generated* *A screen shot of a computer

Description automatically generated*

*Based on the 95% prediction interval for an individual response, we can be 95% certain that the car’s fuel efficiency will fall within the interval of (12.6449, 26.1045).*

*The 95% confidence interval for the mean tells us that we can be 95% certain that the average fuel efficiency for multiple cars that meet the same rear axle ratio and horsepower above will fall within the interval of (17.5716, 21.1777).*

*The prediction interval is wider than the confidence interval since the prediction interval considers the uncertainty that relates to Y as well as sampling uncertainty that is related to the regression parameters (zyBooks, a Wiley brand, 2016).*

*.*

## Conclusion

*I would recommend using this model based on the analysis that I have performed with the assumption that the sample size is a sufficient size. The analysis has shown that both rear axle ratio and horsepower influence fuel efficiency. The analysis also shows an accurate prediction of miles per gallon based on our two predictor variables. The result from the analysis shows there is a strong correlation between rear axle ratio and horsepower on a cars fuel efficiency based on the value is 0.7412. the conclusion of the overall F-test and the individual t-tests that were run also show that the predictor variables are significant and have a relation to the response variable. The practical importance of this analysis is that it can be used to help car manufacturers to estimate the fuel efficiency of cars currently being made as well as to design cars in a way to get better fuel efficiency.*

## Citations

zyBooks, a Wiley brand. (2016). *Applied Regression Analysis (R)*. ZyBooks. https://www.zybooks.com/catalog/applied-regression-analysis-r/