# MAT 303 Module One Problem Set Report

Multiple Regression

Cierra Smith

Cierrs.smith3@snhu.edu

Southern New Hampshire University

## Introduction

I am exploring data for a car maker and looking into the fuel economy. The results from my research can be used so that manufacturers and dealerships can give accurate information on fuel economy for various cars. The results can be also used by consumers to make an educated decision on what car to buy based on fuel economy. I will be using multiple regression and correlation in this problem set. I will generate a QQ plot to test normality and I will generate a residual against fitted plot to test homoscedasticity.

## Data Preparation

*I have a total of 12 variables that I can use; however, I believe the most important variables for this problem set are miles per gallon, quarter mile time and weight. In this data set I have 12 columns and 32 rows.*

## Multiple Regression Model

### Correlation Analysis

*Chart, scatter chart

Description automatically generated*

*Chart, scatter chart

Description automatically generated*

*The relationship between Fuel economy and Weight is that they have a negative correlation. As weight increases the fuel economy decreases. The correlation coefficient is -0.8677 and this is a strong negative correlation. The relationship between fuel economy and horsepower is a negative correlation. As the horsepower increases the fuel economy decreases. . The correlation coefficient is* -0.7762 *and this is a moderate negative correlation.*

### Reporting Results

The general formula for multiple regression is : A screenshot of a computer

Description automatically generated

the general form of the multiple regression model for fuel efficiency (miles per gallon) as the response variable and weight and horsepower as predictors is : The value for R-Squared is 0.8268. 82.68% of the variance in mpg can be explained by weight and horsepower. The value for Adjusted R-Squared is 0.8148. Adjusted R-Squared allows you to adjust the statistics based on the number of independent variables. The beta estimate for weight is and the beta estimate for horsepower is . As the predictor variable increases the response variable decreases based on the beta values. The fitted value is the mean response variables predicted when you include predictor values. Residual values estimate regression error based on the sample multiple regression functions.

*A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated*

Based on the plots the plots meet the conditions for the assumption of homoscedasticity because the residual v fitted value graph has no major clustering of values. Based on the QQ plot it shows normalcy because the points do not deviate from the line to much ( assumption of normality).

### Evaluating Model Significance

*The p value is* 9.109e-12 . 9.109e-12 > 0.05 therefore you reject the null hypothesis. A significant relationship exist between mpg and at least one of the predictor variables. The p value horsepower is 0.00145. 0.00145 > 0.05. *The p value weigh is* 1.12e-6 . 1.12e-6 > 0.05 Therefore you reject the null hypothesis. So, weight and horsepower are important when calculating mpg.

### Making Predictions Using the Model

is the formula that I will use to predict fuel efficiency for a car hat has a wt = 2.95 and a hp = 179. The residual from this observation is as followed . The *95% prediction interval for the car identified in the previous question is (14.645,25.5556). We can predict that 95% of the cars mpg will fall between this interval. The mpg from the previous equation fits within this set of numbers. The 95% confidence interval for the car identified in the previous question is (18.8249,21.3758). I can be 95% confident that the mean mpg’s of a group of car will fall within this interval.*

A screenshot of a computer

Description automatically generated

The prediction is wider than the confidence interval because the prediction interval gives more room for error due to the uncertainty of y varying according to regression error.

## Conclusion

The sample size is large enough for this model. The model has accuracy when representing how the change in horsepower and weight can change mpg. The higher the weight the lower the mpg and the same goes for horsepower. This model depicts a strong negative correlation between weight and horsepower against mpg. This model is practical because it will help car makers predict mpg before they produce a car. Also, car buyers can check mpg’s before deciding what car to buy. This makes buyers more educated in their car buying choices and it helps manufacturers compete against each other in the mpg category.