# MAT 303 Module One Problem Set Report

Multiple Regression

Camilo Hoyos

camilo.hoyos@snhu.edu

Southern New Hampshire University

## Introduction

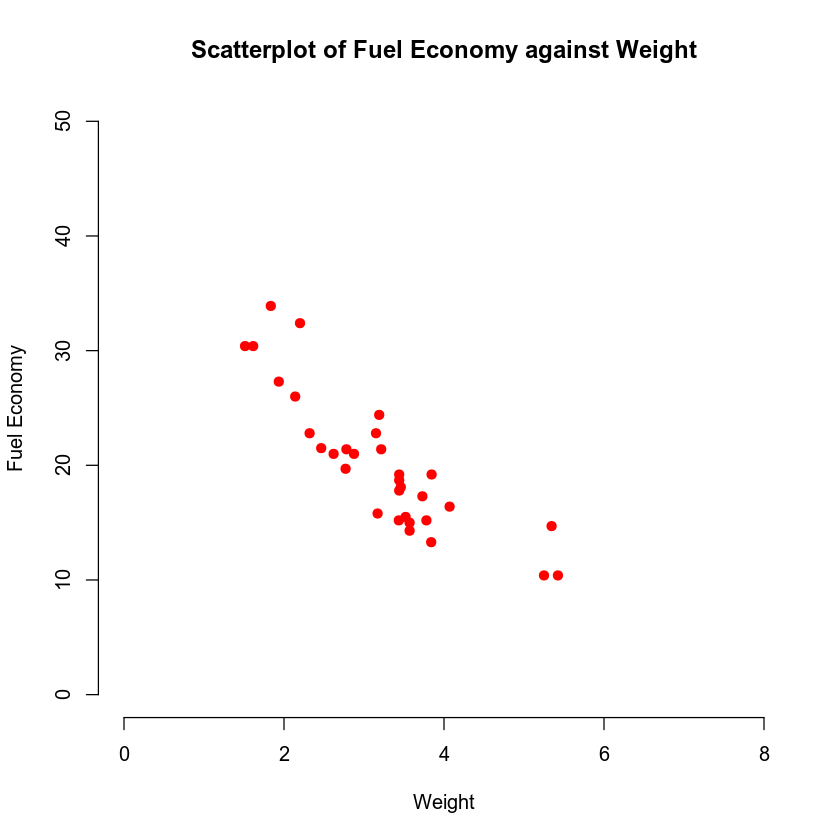
The mtcars data set contains several vehicle attributes such as: miles per gallon (mpg), number of cylinders (cyl), gross horsepower (hp), weight (wt), and more. The results might be used to determine how weight or horsepower correlates to miles per gallon. If there is a desire to include features in a vehicle that impact the weight of the vehicle, one can identify how much their MPG might be impacted. Or if one may want to increase the horsepower of a vehicle for market differentiation, one can identify how much their MPG might be impacted. I will be running a multiple regression model using mpg as the dependent variable, and weight and horsepower as the two dependent variables.

## Data Preparation

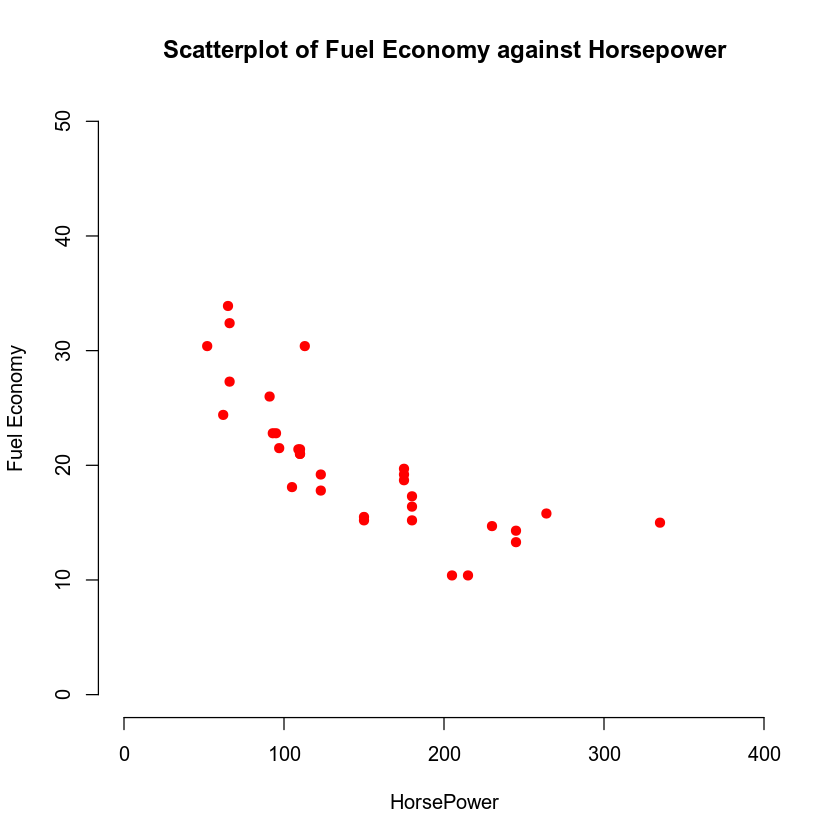
There are three important variables in this data set. Miles per gallon (mpg) is the dependent variable. Weight and horsepower are the two independent variables. We will be identifying what the correlation between weight and horsepower is to miles per gallon. There are a total of 12 columns but we will only use the three relevant columns. There are 32 rows in this data set.

## Multiple Regression Model

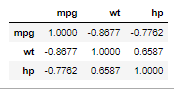
### Correlation Analysis

**

The relationship between fuel economy and weight is that fuel economy drops as weight increases.

**

The relationship between fuel economy and horsepower is that fuel economy drops as horsepower increases.

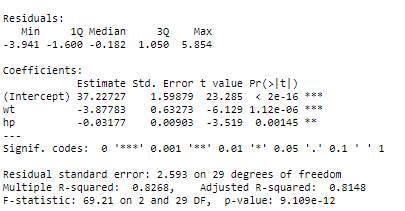
**

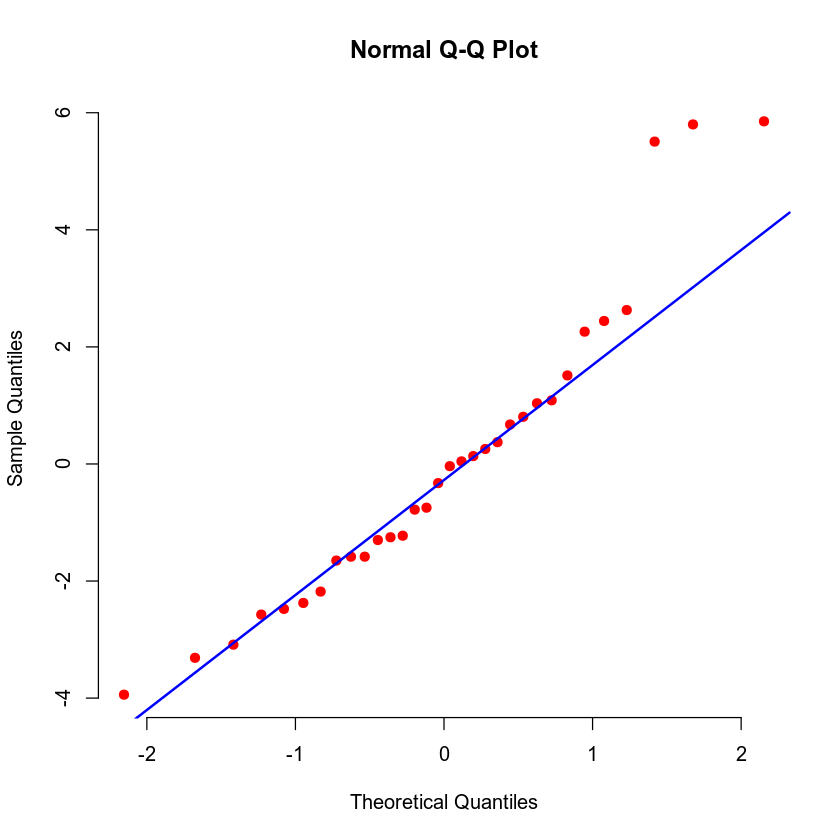
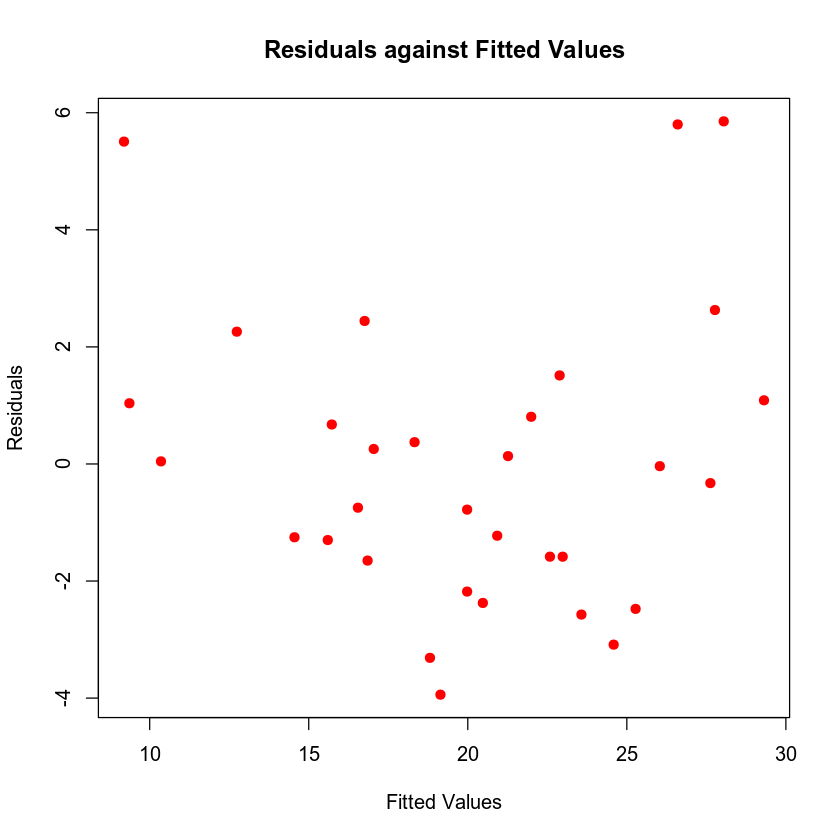
What this correlation coefficient tells us how mpg is impacted based on the relationship with weight and horsepower. What this provides is that there is a very strong negative correlation between miles per gallon and weight. There is also a very strong negative correlation between miles per gallon and horsepower.

### Reporting Results

The multiple regression model is where Y = Miles per Gallon, X1 = Weight, and X2 = Horsepower. Y is the response variable and X1 and X2 are the predictor variables.

The multiple regression model for fuel efficiency is . The coefficient of multiple determination is 0.8268. What this provides is that 82.68% of the variance in fuel efficiency can be explained by weight and horsepower. The adjusted coefficient of multiple determination is 0.8148 or 81.48%. These values are within the range of multicollinearity. I would say that this is justified as well since comprehensive engine systems that output more horsepower generally weigh more. The beta estimate for weight is -3.87783 and the beta estimate for horsepower is -0.03177. What this describes is that for every unit of weight that increases fuel efficiency will drop 3.87783 units, and that for every unit of horsepower that increases fuel efficiency will drop 0.03177 units. A fitted value is the predicted value of Y for the ***i***th sample of the predictor values where ***i*** is a selected incremented value. A residual is the estimated regression error based on the sample multiple regression function. Residuals against fitted values and the Normal Q-Q plot are below. (Berrier et al., 2016)

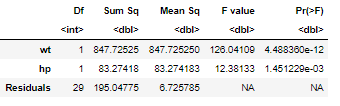
**



What we can interpret from the residuals against fitted values is that this does not violate the constant variance assumption. Residuals of each set of values for the predictor variables should have equal or similar variance. This is known as homoscedasticity. Since we see no evidence of nonlinearity. The Q-Q plot helps with identifying the Normality assumption. That is the residuals of each set of values for the predictor variables form a normal distribution. (Berrier et al., 2016)

### Evaluating Model Significance

The overall F-test identifies whether a linear relationship exists with at least one predictor variable. H0 would be the null hypothesis that is not useful because the slope of all parameters are equal to 0. If Ha has a predictor variable that does provide a non-zero slope then we have an alternative hypothesis. The P-value calculated is 9.109e-12 which is lower than a 5% level of significance. With the p-value and seeing that F = 69.21 we can reject the H0.

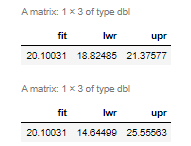


From the analysis of variance we can see that the P-value for each of the predictor variables is lower than a 5% level of significance. The 95% confidence intervals have a fit of 23.57233, a lower limit of 22.45623 and an upper limit of 24.68843. 95% of the values of this sample should exist in this range.

### Making Predictions Using the Model

*Make predictions using the regression model. Address the following questions in your analysis:*

The predicted value of a car that weighs 2950 lbs and a horsepower of 179 is 20.10031. The actual value is 22.7 miles per gallon. The residual for this is 2.59969. The 95% prediction interval has a fit of 20.10031, a lower limit of 14.64499 and an upper limit of 25.55563. The 95% confidence interval has a fit of 20.10031, a lower limit of 18.82485 and an upper limit of 21.37577. Again this simply states what the upper and lower limit of 95% of the sample will land. The distinction is that a prediction interval takes uncertainty according to the regression error. (Berrier et al., 2016)

**

## Conclusion

I would recommend using this model since it meets a general size of at least 30 records. (Berrier et al., 2016) We have also verified mean of zero, normality, constant variance, and have identified a potential form of multicollinearity. If the correlation between weight and hp is expected then this can be accepted.

What we have evaluated is that there is a negative correlation between weight and fuel efficiency and horsepower and efficiency. That for every unit of weight and horsepower, it will reduce a vehicle's fuel efficiency. We can confirm this hypothesis since we have a p-value of 9.102e-12 which is well below the level of significance that we can accept the alternative hypothesis. With the assumptions made in multiple regression analysis we can consider this a valid model.

The practical importance is that a prediction can be made based on a vehicle being manufactured based on its weight and horsepower. A predicted value can be identified using the multiple regression model . When the vehicle is finally measured, another point can be added in the model to further refine and predict a future vehicle’s fuel economy based on weight and horsepower.

## Citations

Berrier, J., Pardoe, I., Sturdivant, R. X., & Watts, K. (2016). *Applied Regression Analysis (R)*. Zyante Inc.