# MAT 303 Module One Problem Set Report

Multiple Regression

[Luna Saccoia]

[Luna.Saccoia@snhu.edu]

Southern New Hampshire University

## Introduction

*The statistical analyses being performed here have to do with calculating the relationship between a vehicle’s weight and its fuel efficiency, as well as its horsepower and fuel efficiency.*

*The data set I am exploring is weight and horsepower in how they each relate to fuel efficiency. The calculations and models should tell us if a car’s weight affects its miles per gallon as well as if its horsepower affects its miles per gallon.*

*These results might be used by car manufacturers to see if trimming weight or lowering horsepower on a car will help it get better fuel economy. This can be useful if they are able to trim weight and horsepower to improve fuel economy without sacrificing too much power. If they can find a happy medium, people will be more drawn to the vehicle. Gas is expensive, so people are looking to have vehicles that get better gas mileage, but people also want a car that has enough power for their needs. This is why finding that happy medium is key for car manufacturers when making vehicles.*

*The analyses we will be doing here are to create a multiple regression model, find the fitted values, find the residuals, plot the fitted values against the residuals, create a q-q plot to test assumptions of normality of the residuals, and find the confidence interval. We will take all of this information and use it to draw conclusions about whether or not weight and/or horsepower have a statistically significant effect on gas consumption in miles per gallon.*

## Data Preparation

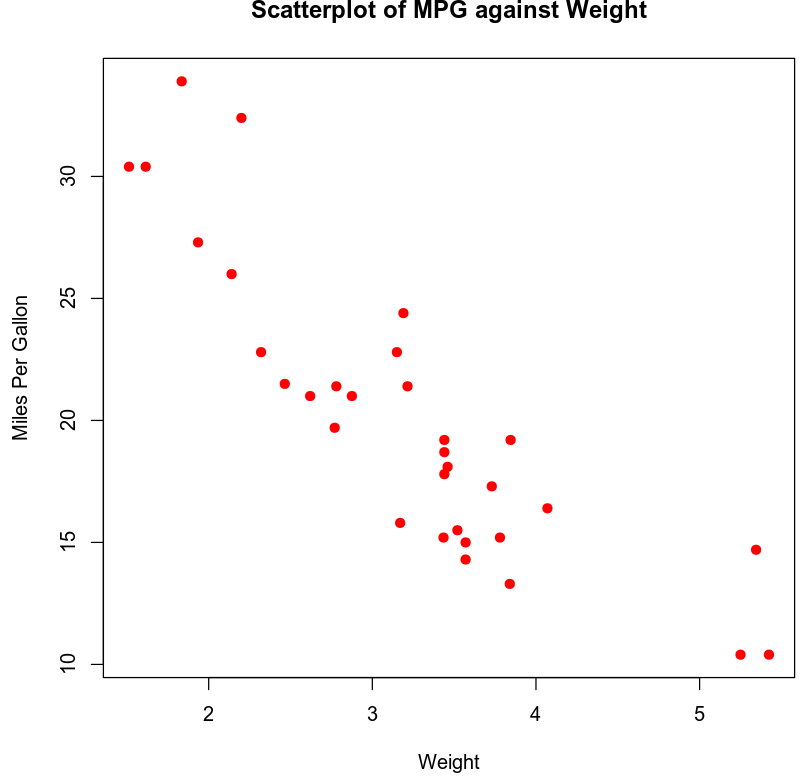
*The important variables I will be analyzing are weight, horsepower and miles per gallon. There are 12 columns in the full data set, for different variables relating to the vehicles. There are 32 rows, one for each vehicle whose data is included in the data set.*

## Multiple Regression Model

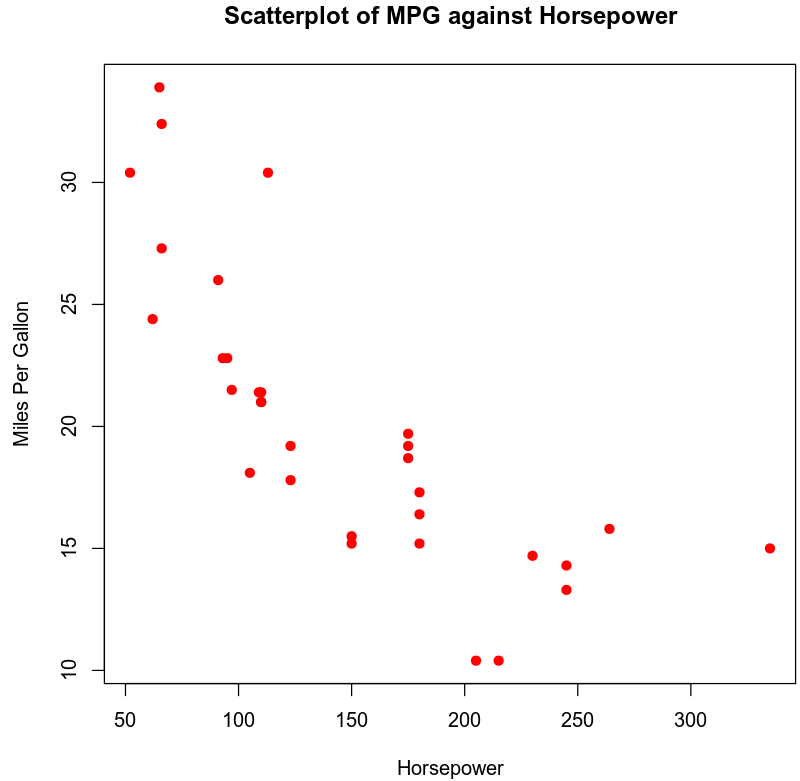
### Correlation Analysis

*We used scatterplots to visualize the relationship between the variables contained within the data set, which were relevant to our model. These variables, whose relationships are relevant here, are weight and horsepower, both tested against miles per gallon.*

*Below is the scatterplot of fuel efficiency (miles per gallon) against weight. It shows a negative correlation. I did run the analysis with just weight against miles per gallon, to accurately report how strong the correlation is. It is considered a moderate correlation, due to the adjusted being 0.7446, which is slightly less than 0.80. We see from the scatterplot that as weight increases, fuel efficiency decreases.*

**

*Next is the scatterplot of fuel efficiency (miles per gallon) against horsepower, included below. It also has a negative correlation. Testing just horsepower against miles per gallon gives us an adjusted of 0.589 which shows a moderate correlation. We see from the scatterplot that as horsepower increases, fuel efficiency decreases.*

**

*Next we calculated the correlation coefficients between fuel efficiency (miles per gallon) and weight and between fuel efficiency and horsepower. Our correlation coefficient for weight is -0.8676594 and for horsepower the correlation coefficient is -0.7761684. This means there is a strong negative correlation between weight and MPG, and a moderate negative correlation between horsepower and MPG.*

### Reporting Results

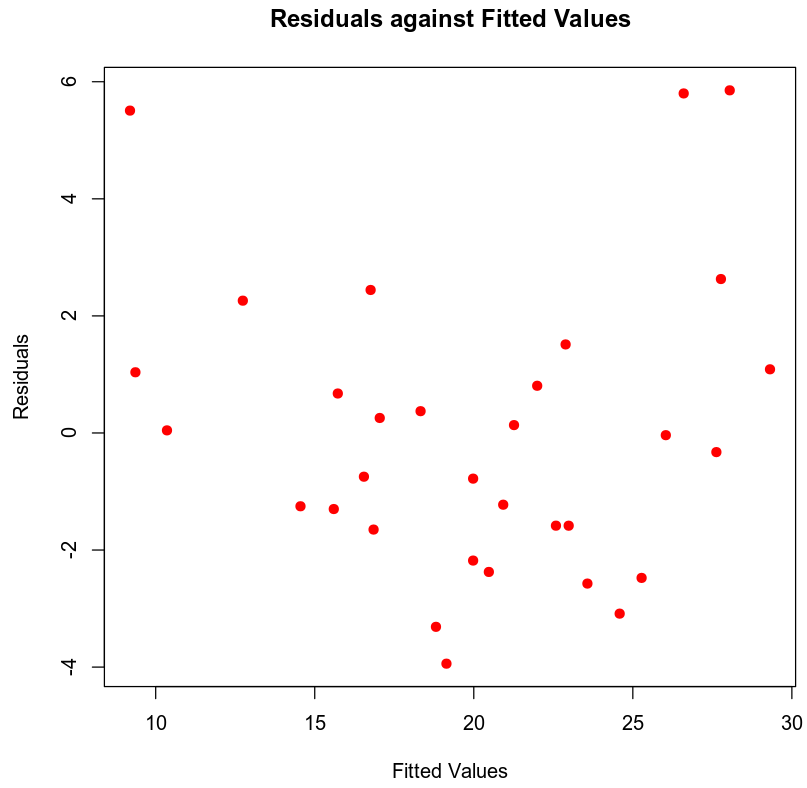
*The equation for the multiple regression model for fuel efficiency as the response variable and weight and horsepower as the predictor variables is:*

*Using this equation, we can input values from our data set. We can choose a vehicle from our data set and input that vehicles weight into our equation for and horsepower for . Doing this would allow us to calculate the predicted fuel efficiency of the chosen vehicle.*

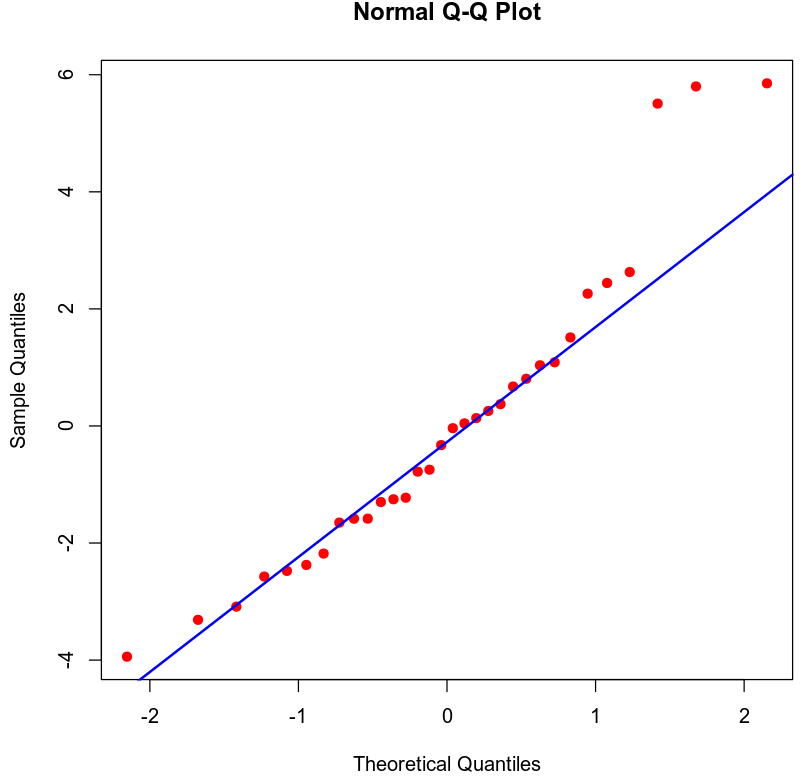
*The (R-squared) value is 0.8268 and the (Adjusted R-squared) value is 0.8148. This tells us that approximately 83% of the variation in fuel economy can be explained by a model that uses predictors for weight and horsepower.*

*The beta estimates are and . This means that for each 1 unit increase in our predictor variable, our response variables will decrease by 3.87783 and 0.03177 respectively.*

*The fitted values and residuals were obtained and a scatterplot is below for the Residuals against Fitted Values.*

**

*Below is the Normal Q-Q plot.*

**

*Based on these plots we see the plot for residuals against fitted values meets the requirements for homoscedasticity. There is no discernible clustering or pattern to it. I would conclude that the q-q plot has non-normally distributed residuals, as we see that some of the points deviate significantly from the line.*

### Evaluating Model Significance

*For evaluating model significance for the multiple regression model, we will be looking to see if the model is significant at a 5% level of significance. The equation for our overall F-test is:*

*Our P-value is 9.109e-12 which is significantly lower than our level of significance of 0.05. Based on this we should reject our null hypothesis in favor of our alternative hypothesis. This means that a significant linear relationship exists between fuel efficiency and at least one of our response variables (weight, horsepower).*

*For evaluating for each variable to see if each one is significant individually at a 5% level of significance, we use an individual t-test. The equation for this is:*

*Our p-value for weight is 1.12e-06 and our p-value for horsepower is 0.00145. Both of these are significantly smaller than our level of significance of 0.05. Based on these findings, we would reject our null hypothesis in favor of our alternative hypothesis. This test tells us that both weight and horsepower individually have a statistically significant relationship with fuel efficiency, as well as when evaluating with both variables.*

### Making Predictions Using the Model

*The equation to predict fuel efficiency for a car that has a weight of 2.95 and a horsepower of 179 is:*

*If this car achieves an average of 22.7 miles per gallon, this would be our sample observation denoted by*

*To find our residual, we use the equation So . So our residual is 2.5997.*

*The 95% prediction interval for an individual response for fuel economy of this car is .*

*The prediction interval for an individual response tells us we can be 95% certain that a car’s fuel economy (in miles per gallon) will fall within these lower and upper bounds, if the car’s weight is 2.95 and its horsepower is 179. This is taking into account uncertainty related to  varying according to the regression error, , as well as sampling uncertainty related to estimating the regression parameters.*

*Taking the regression error into account as well as sampling uncertainty relating to estimating the regression parameters is the reason the prediction interval for an individual response is wider than the confidence interval for the mean. (Zybooks, 2016)*

*The 95% Confidence interval for fuel economy of this car is .*

*This confidence interval for the mean tells us we can be 95% certain that the average fuel economy (in miles per gallon) of a group of cars will be within these lower and upper bounds; if they have a weight of 2.95 and a horsepower of 179.*

## Conclusion

*Based on the analysis I have performed I would use this model, as long as the sample size was sufficiently large. It shows accurately that horsepower and weight effect fuel efficiency. Based on the analysis we can conclude that the heavier a car is and the higher its horsepower, the lower its fuel efficiency will be.*

*These results show that there is a strong correlation between weight and horsepower on fuel efficiency. This is found by calculating our value. When using both response variables we have an value higher than 0.80, which tells us the correlation is considered strong.*

*The practical importance of this analysis is that it helps car manufacturers predict how fuel efficient vehicles in the design phase will be. They can use the weight of the vehicle and the horsepower it will have to calculate how fuel efficient it will be. This will be useful to know so they do not waste time making a car that will not be appealing to their customers. They will be able to use this information and predictions to make changes while still in the design phase, so when they get to the production phase, changes are already made and they can save time and money by only sending appealing cars to the production phase.*

## Citations

Zybooks MAT 303: Applied Statistics II for Science, (2016, August).

Retrieved March 7, 2020, from https://learn.zybooks.com/zybook/SNHUMAT303v1