# MAT 303 Module One Problem Set Report

Multiple Regression

Justin Farquhar

Justin.farquhar@snhu.edu

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## Introduction

We are utilizing a data set containing the fuel economy of cars containing several important variables that are associated with fuel economy. These results could be use to help find correlations between various variables to see how they effect fuel economy, either on their own or in conjunction with one another. We will be performing analyses on some of the various variables in correlation with mpg, seeing how the variables correlate in different graph types allowing us to better see the data.

## Data Preparation

I believe some of the most important variables in this data include mpg, hp, wt, and qsec. These are a bulk of the variables that give us a better gauge of what is going on. For mpg, or miles per gallon, this is the variable we are going to be constantly comparing everything to, seeing how it changes based on the other variables. For hp, or horsepower, this shows how strong the engine is and how quickly it is able to move, which I imagine will have some surprising effects on the fuel economy of the vehicle. For wt, or weight, the heavier the vehicle, the less efficient it is going to be as it is going to take much more power to move the entire vehicle. For qsec, or the quarter mile time, this gives a measure of the acceleration of the vehicle, which should mostly correlate with the horsepower. The other variables are still important but are likely to have less of an effect on mpg as a whole. There are a total of 32 rows and 12 columns throughout this data set.

## Multiple Regression Model

### Correlation Analysis

*A graph with red dots

Description automatically generated*

Based on this scatter plot, rear axel ratio appears to have a positive relationship with fuel economy. As the ratio increases, so does the fuel economy of the vehicle.

A graph with numbers and blue dots

Description automatically generated

This next graph is comparing the effect horse power has on the fuel economy. Looking at the graph, it is clear to see that there is a negative relationship between miles per gallon and horse power, meaning that as horse power increases, the fuel economy continues to decrease. The data does seem to taper off at the end between the last few points which is likely due to a baseline efficiency that manufacturers are supposed to be hitting, as if efficiency gets too low, there would be an extremely limited population willing to purchase the vehicle.

**A screenshot of a computer

Description automatically generated**

Looking at the results of this table showing the correlation matrix, you can see that the correlation between mpg and hp is -0.7762, which shows that there is a negative linear relationship between the two. The closer to -1 the number is, the stronger the relationship is, eventually leading to a perfect linear relationship, so a relationship of -0.7762 is a fairly strong relationship. To then move onto the relationship between drat and mpg, you can see that it is 0.6812 which indicates a positive linear relationship. Similarly, the negative linear relationships, the closer this value is to 1, the stronger the relationship between the values. In this case, 0.6812 indicates that the relationship is fairly strong positive relationship.

### Reporting Results

General Form:

Prediction Model:

Prediction model equation using outputs from multiple regression model:

For this data, my value of R-squared is 0.7412 and the adjusted R-squared is 0.7233. The R-squared value indicates that roughly 74.12% of the variability in miles per gallons is explained by real axle ratio and horsepower in this model. The remaining percentage would be accounted by the other variables that we did not account for presently. The adjusted R-squared being relatively close to R-squared indicates that the predictors that we added to the model are relevant and provide meaningful information.

The estimated coefficient for variable rear axle ratio is 4.6982. Meaning that on average, the fuel economy of the car increases by 4.6982 for each unit increase in the rear axle ratio. In the same sense, for the estimated coefficient for variable horsepower it is -0.0518. Meaning that on average, the fuel economy of the car decreases by 0.0518 for each unit increase in horsepower. This may not seem like much of an adjustment, but with horsepower starting around 80 in this data set and getting up to over 350, increases can add up quickly and drop the fuel efficiency of the vehicle.

A fitted value is the estimated value of the response variable (mpg in this case) based on the regression model. The fitted value is what the model predicts mpg will be given the values of the predictors (rear axle ratio and horsepower in this case). A residual is the difference between the observed value of the response variable and the fitted value for each point. It ultimately shows how much the prediction of the model is off from the actual observed value.

A graph with green dots

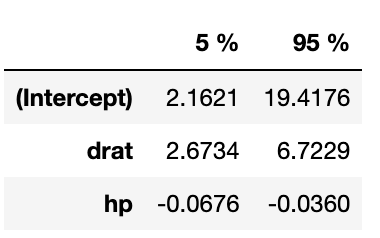
Description automatically generated

A graph with red dots

Description automatically generated

When we consider homoscedasticity we are trying to ensure that the residuals of the model have a constant variance across all levels of the predictor variables, to where no matter what values of rear axle ratio and horsepower the spread of the residuals should stay roughly the same. Based on the plot point, the data mostly complies to homoscedasticity, though there are some groupings leading to some slight heteroscedasticity, but this is minor and would need more extensive data to confirm. Looking at the normality of the residuals through the Q-Q plot, there is a slight departure from normality in the trails, but this is not directly a major issue and again would need some additional data to confirm or deny, though caution should be used until that further data is tested.

### Evaluating Model Significance

**

*A black text on a white background

Description automatically generated*

If we specifically look at the Pr(>F) which is the P-value for the equation, it shows that each values for rear axle ratio and horsepower are less than 0.05 or 5%, suggesting that each predictors are significant on their own. To consider the null hypothesis, each of the coefficients for the predictors would be zero leading to there being no relationship between the response variable and the predictors. For the alternative hypothesis, at least one of the predictors has a non-zero coefficient, indicating a significant relationship with the response variable. Seeing as the P-values are well below 0.05, we can reject the null hypothesis. The overall model is significant at the 5% significance level.

### 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 would plug it into the prediction equation:

The residual for this situation if the car achieves an average of 20.5 miles per gallon would be 1.1268.

A screenshot of a cell phone

Description automatically generated

Based on this screenshot, we can see that the model predicts the fuel efficiency with the same value through the prediction equation of 19.3747 and it also gives an upper and lower value for this prediction. Looking at these values, we are 95% confident that the individual vehicles mpg will be between this lower and upper value. This interval also accounts for the natural variability in individual observations. The same thing is true then for the confidence interval, it predicts the value again, while also providing a lower and upper range that has 95% confidence that the true mean mpg will be within. The prediction interval is wider due to it accounting for two sources of variability, the estimating the mean and the natural variability of individual observations around the mean.

## Conclusion

*Describe the results of the statistical analyses and address the following questions:*

Based on the analysis performed, and after increasing the sample size of the data, I would recommend using this model with some considerations. The model includes rear axle ratio and horsepower as predictors of fuel efficiency, both of which show significant relationships. The high statistical significance of these predictors shown by very low p-values and the model’s overall significance shown by the F-test suggest that these variables do contribute meaningful information about fuel efficiency. As for the considerations, as more data is added, we will need to ensure that the model fit or R-squared and adjusted R-squared values are still reasonably high indicating that the model explains a substantial proportion of the variability in mpg, making it more reliable. We will also need to ensure that the assumptions of regression still hold, those of which being: linearity, independence, homoscedasticity, and normality of residuals. In statistical terms, the coefficients hold up to where for each unit of increase of rear axle ratio, the fuel efficiency increases by 4.6982 and for each unit increase of horsepower decreases the fuel efficiency by 0.0518. The practical importance of these analyses can be extremely helpful when considering production for future vehicles. Being able to pinpoint a specific rear axle ratio and horsepower to deliver better fuel efficiency. This could also be a great starting point for further studies. Continuing to see what other factors have effects on fuel efficiency could lead to breakthroughs in the industry, potentially drastically improving fuel efficiency.