# MAT 303 Module One Problem Set Report

Multiple Regression

Brandon B. Gibbs

brandon.gibbs@sbhu.edu

Southern New Hampshire University

## Introduction

This study conducts statistical analyses to explore the relationships among a vehicle's fuel efficiency (measured in miles per gallon), rear axle ratio (drat), and horsepower (hp). The dataset includes variables such as vehicle weight and hp, examining their impacts on fuel efficiency.

The primary objective is to determine the influence of drat and hp on mpg using multiple regression models. Results are anticipated to yield insights beneficial for automotive manufacturers aiming to enhance fuel efficiency without compromising performance.

As fuel prices rise, consumers increasingly seek vehicles with high gas mileage and sufficient power output. Thus, finding an optimal balance between weight reduction and hp modification is essential for manufacturers targeting eco-conscious buyers.

The analytical framework involves constructing a multiple regression model, calculating fitted values, determining residuals, and plotting fitted values against these residuals. A Q-Q plot will assess residual normality, alongside the establishment of confidence intervals. This statistical data will help evaluate the significance of drat and hp on fuel efficiency (mpg).

## Data Preparation

The dataset under consideration, known as mtcars, comprises 32 observations and 12 variables, encompassing a range of car attributes that are significant for our analysis. The primary variables of interest include fuel efficiency (measured in miles per gallon, mpg), number of cylinders (cyl), engine displacement (disp), horsepower (hp), rear axle ratio (drat), vehicle weight (wt), quarter-mile time (qsec), engine shape (vs), type of transmission (am), number of gears (gear), and number of carburetors (carb). These variables collectively provide a robust framework for evaluating various car models and their specifications, thereby facilitating an in-depth exploration of the relationships between fuel efficiency and other relevant automotive characteristics.

For the purpose of this analysis, the focus will be primarily on fuel efficiency (mpg), rear axle ratio (drat), and horsepower (hp). A thorough data preparation process is crucial for ensuring that the statistical analyses conducted are both accurate and meaningful. This preparation entails several steps, including the loading and initial inspection of the dataset, management of any missing values, normalization of data via scaling, and the subsetting of data tailored for analysis. The analytical procedures will be executed utilizing Jupyter Notebooks in conjunction with R programming language, which will allow for a comprehensive assessment of the relationships present within the data.

## Multiple Regression Model

Correlation analysis is a statistical method employed to assess the strength and direction of the relationship between two variables. In our study, we will utilize correlation analysis to investigate the relationship between fuel efficiency (measured in miles per gallon, or mpg) and horsepower (hp). To visualize these correlations, we will employ scatterplots, which will provide a clear and comprehensive understanding of the data. This approach will help identify potential predictors and ensure the accuracy of our regression models. Subsequently, we will construct the correlation matrix for our coefficients.

A graph with purple dots

Description automatically generated

The scatterplot illustrates a positive correlation between fuel efficiency (mpg) and the rear axle ratio (drat) meaning that as the rear axle ratio increases, the fuel efficiency of the vehicles tends to improve. However, the data points are somewhat dispersed, suggesting that while there is a general upward trend, the relationship is not particularly strong. This dispersion indicates that other factors might also be influencing fuel efficiency, and the rear axle ratio alone does not fully explain the variations in mpg.

**A graph with purple dots

Description automatically generated**

The scatterplot shows a clear inverse relationship between fuel efficiency in Miles Per Gallon (mpg) and horsepower (hp). It shows that as a vehicle's horsepower increases, its fuel efficiency tends to decrease. The downward slope of the data points visually represents this trend. Normally maintained vehicles with higher horsepower generally require more fuel to produce greater power, which results in lower fuel efficiency.

The Pearson Correlation Coefficient between fuel efficiency (mpg) and rear axle ratio (drat) is 0.6812. This value indicates a strong positive correlation between these two variables. In other words, as the rear axle ratio increases, fuel efficiency also tends to increase. This positive relationship suggests that vehicles with higher rear axle ratios generally achieve better fuel efficiency.

On the other hand, the Pearson Correlation Coefficient between fuel efficiency (mpg) and horsepower (hp) is -0.7762. This value signifies a strong negative correlation between these variables. This means that as horsepower increases, fuel efficiency tends to decrease. The negative relationship indicates that vehicles with higher horsepower typically consume more fuel, resulting in lower fuel efficiency. This is consistent with the understanding that more powerful engines require more fuel to operate, thereby reducing overall fuel efficiency.

### Reporting Results

**Interpretation of Beta Estimate**

From the output the prediction model equation



The intercept value of 10.7899 signifies that, theoretically, when both the rear axle ratio (drat) and horsepower (hp) are equal to zero, the expected miles per gallon (mpg) is 10.7899. However, this value lacks practical significance, as it is not feasible for either drat or hp to be zero in actual vehicles. The coefficient for drat, quantified at 4.6982, indicates that an increase of one unit in the rear axle ratio is associated with an expected increase of 4.6982 in mpg, assuming that horsepower remains constant. This suggests a positive correlation where vehicles equipped with higher rear axle ratios tend to exhibit improved fuel efficiency.

Conversely, the coefficient for hp, recorded as -0.0518, suggests that for each unit increase in horsepower, the expected mpg decreases by 0.0518, provided that the rear axle ratio remains constant. This finding reflects a negative correlation, indicating that vehicles with greater horsepower generally exhibit lower fuel efficiency.

The residual standard error is measured at 3.17 with 29 degrees of freedom, which quantifies the typical magnitude of prediction errors. This statistic provides insight into the extent to which actual mpg values deviate from the predicted values derived from the model. The multiple R-squared value of 0.7412 indicates that approximately 74.12% of the variability in fuel efficiency (mpg) can be accounted for by the model. This suggests a strong relationship between the predictor variables (drat and hp) and the response variable (mpg).

Furthermore, the adjusted R-squared value of 0.7233 refines the R-squared value by taking into account the number of predictor variables included in the model, thereby offering a more precise representation of the model's explanatory capacity.

Lastly, the F-statistic value of 41.52, with 2 and 29 degrees of freedom, alongside a p-value of 3.081e-09, strongly supports the statistical significance of the model. This indicates substantial evidence that at least one of the predictor variables (drat or hp) demonstrates a significant relationship with the response variable (mpg).

A graph with purple dots

Description automatically generated

The "Residuals against Fitted Values" plot illustrates an absence of discernible patterns, which suggests that the variance of the residuals remains consistent across different levels of fitted values. This observation supports the assumption of homoscedasticity, as the residuals are uniformly distributed around the horizontal axis. Such a distribution indicates that the variability of the residuals does not change with the magnitude of the fitted values, thereby reinforcing the validity of the model's assumptions in this context.

A graph of a normal q-q plot

Description automatically generated

The “Normal Q-Q Plot” compares the quantiles of the residuals with theoretical quantiles from a normal distribution. Most points lie close to the diagonal line, indicating that the residuals are normally distributed. However, there are some deviations at the tails, suggesting slight departures from normality. Overall, the residuals mostly conform to the normality assumption, which is important for valid inference in regression analysis.

### Evaluating Model Significance

To ascertain the model's significance, we examined the P-value derived from the F-test. A P-value less than the alpha threshold of 0.05 leads us to reject the null hypothesis, thereby concluding that the model is statistically significant and that at least some of the predictor variables exert an influence on the outcome variable.

In addition to the overall model assessment, we conducted individual tests on the regression coefficients corresponding to the variables of interest, specifically the rear axle ratio (drat) and horsepower (hp). For each of these coefficients, the null hypothesis (H0) asserts that the coefficient is equal to zero, suggesting no effect, while the alternative hypothesis (H1) posits that the coefficient is not equal to zero, indicating a significant relationship. If the P-value for either variable is less than 0.05, we reject the null hypothesis for that specific coefficient, thereby demonstrating statistical significance and affirming that the variable plays a role in predicting fuel efficiency.

To further interpret the implications of our findings, we constructed 95% confidence intervals for the estimated parameters associated with drat and hp. These confidence intervals provide a statistical range within which we expect the true population parameter values to reside with 95% certainty. This additional analysis enhances our understanding of the magnitude and direction of the relationship each predictor variable has on fuel efficiency, as well as the degree of uncertainty associated with our estimates. Such insights are crucial for both theoretical interpretations and practical applications in the field of fuel efficiency analysis.

### Making Predictions Using the Model

Predicted Fuel Efficiency:

A math equations and numbers

Description automatically generated with medium confidence

For a car with a rear axle ratio (drat) of 3.15 and a horsepower (hp) of 120, the predicted fuel efficiency is 19.3747 miles per gallon (mpg). If the car achieves an average of 20.5 mpg, the observed residual is 1.1253. The predicted interval for the car's fuel efficiency ranges from 13.4531 to 24.9653 mpg, meaning we can be 95% confident that the actual fuel efficiency of a similar car will fall within this range. This interval accounts for variability in individual observations.

The confidence interval ranges from 17.8767 to 20.8726 mpg, indicating that we can be 95% confident that the mean fuel efficiency for cars with similar characteristics (drat = 3.15 and hp = 120) will fall within this range. This interval reflects the precision of the estimated mean response.

The prediction interval is wider than the confidence interval because it considers both the uncertainty in the estimated mean response and the variability of individual observations. In contrast, the confidence interval only accounts for the uncertainty in the estimated mean response. Thus, the prediction interval includes an additional component to capture the variability of individual data points around the mean.

## Conclusion

For a car with a rear axle ratio (drat) of 3.15 and a horsepower (hp) of 120, the predicted fuel efficiency is 19.3747 miles per gallon (mpg). If the car achieves an average of 20.5 mpg, the observed residual would be 1.1253. The prediction interval for the car's fuel efficiency ranges from 13.4531 to 24.9653 mpg, which means we can be 95% confident that the actual fuel efficiency of a similar car will fall within this range. This accounts for variability in individual observations.

The confidence interval ranges from 17.8767 to 20.8726 mpg, indicating that we can be 95% confident that the mean fuel efficiency for cars with similar characteristics (drat = 3.15 and hp = 120) will also fall within this range. This interval reflects the precision of the estimated mean response.

The prediction interval is wider than the confidence interval because it takes into consideration both the uncertainty in the estimated mean response and the variability of individual observations. In contrast, the confidence interval accounts only for the uncertainty in the estimated mean response. Therefore, the prediction interval incorporates an additional component to capture the variability of individual data points around the mean.