

Mandatory Assignment 1

August Høg Dedenroth (Affiliation: KEA, augu0577@stud.kea.dk)
Danije Gitanovicl (Affiliation: KEA, Dagi0001@stud.kea.dk)

April 17, 2024

GitHub Repository

[GitHub Repository](#)

abstract

The project aims to predict cars' MPG (Miles Per Gallon) based on various features and ascertain their fuel efficiency. By examining the correlation between MPG and features, we identified weight, horsepower, and displacement as having the most significant impact on MPG. These variables allow us to predict MPG effectively and judge a car's fuel-efficiency.

1 Introduction

This paper aims to predict car fuel efficiency (in miles per gallon) based on the provided data. Our null hypothesis posits that the features of cars in the dataset have no correlation with miles per gallon. We will challenge this hypothesis by examining the dataset to identify relationships between the data and MPG, and subsequently use our findings to classify fuel-efficient cars.

So in simple terms, we have two problems we want to solve.

- How can we predict the fuel efficiency based on our data's features?
- Based on previous findings, how can we determine if a car is fuel-efficient or not?

2 Methods

We'll employ multiple methods to analyze the datasets, including multiple regression analysis techniques such as OLS (Ordinary Least Squares)[1] and linear regression. These methods will help predict the target variable MPG using other variables like weight, horsepower, etc. Additionally, we've utilized models for logistic regression[4] to determine a car's fuel efficiency based on our earlier predictions disproving the null hypothesis.

3 Analysis

We checked the data for trends, then used regression to see how different factors affect MPG. We measured how well our models fit the data and looked at residuals to make sure they were good. This helped us learn more about car fuel efficiency, judging whether a car was fuel-efficient based off what factors affected MPG the most.

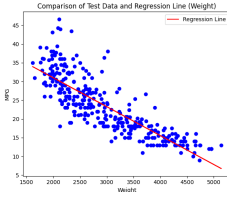


Figure 1: Regression Plot

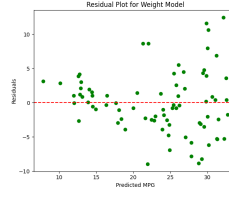


Figure 2: Residual Plot

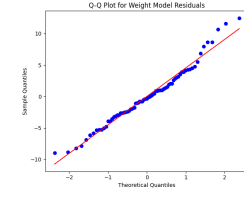


Figure 3: QQ Plot

4 Findings

Our analysis shows that weight, horsepower, and displacement significantly influence MPG. This is evident in both the plots and summary data, where displacement has a coefficient of 0.0174, horsepower -0.0103, and weight -0.0068. The regression models demonstrate moderate to high accuracy in predicting MPG, with slight variations depending on the chosen MPG data. The regression line closely matches the actual data, although some points deviate slightly, indicating our predictions reliably estimate MPG. Residual analysis reveals most points clustered around 0, with some scattered further away, suggesting overall accuracy in our model's predictions.[3] When looking at the Q-Q plot we can see that the data closely matches a normal distribution. The other two Q-Q plots are somewhat similar to this one.[2]

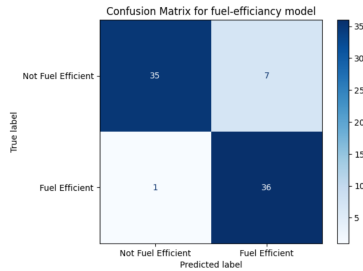


Figure 4: Regression Plot

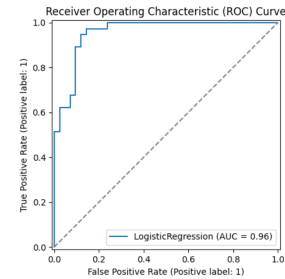


Figure 5: Residual Plot

Regarding fuel efficiency, we can ascertain a car's status based on weight, horsepower, and displacement due to their significant relationships. We developed a model to determine car fuel efficiency using the MPG median-threshold and relevant features like weight, displacement, and horsepower. While the matrix reveals minimal false-positive and false-negative cases, they remain within an acceptable range. This is reinforced by the ROC curve, illustrating the trade-off between true and false positive rates at various thresholds.

5 Conclusion

- How can we predict the fuel efficiency based on our data's features?

In summary, our research reveals how various car attributes, like weight, horsepower, and displacement have a large impact on fuel efficiency. While weight seems to have the most impact, horsepower and displacement seems to also contribute to an extent. Using regression analysis, we've gained valuable insights into what drives fuel efficiency in cars.

- Based on previous findings, how can we determine if a car is fuel-efficient or not?

We have now also defined a binary-classification for what determines if a car is fuel-efficient or not and trained a model to recognize it.

References

- [1] Burton, A. L. (2021). OLS (Linear) Regression. The Encyclopedia of Research Methods in Criminology and Criminal Justice
- [2] nalyticsvidhya.com - Akansha Khandelwal (06 Sep, 2021). How Q-Q plots can help us identify the distribution types?
- [3] towardsdatascience.com - Usman Gohar (Mar 5, 2020). How to use Residual Plots for regression model validation?
- [4] Davis, J., and Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning.