cars

Multiple Regression

Introduction

mtcars is a data set that includes several important variables that are associated with fuel economy. The data will be used to create multiple regressions models to predict the factors strongly correlated with better fuel economy. There are 12 columns and 32 rows included in this data set. Listed below are the important variables with their description. These variables will be used to build the models and perform the analysis.

drat

Rear axle ratio indicates the number of turns of the drive shaft for every one rotation of the wheel axle. A vehicle with a high ratio would provide more torque.

Mpg

Fuel economy in miles/(US) gallon

hp

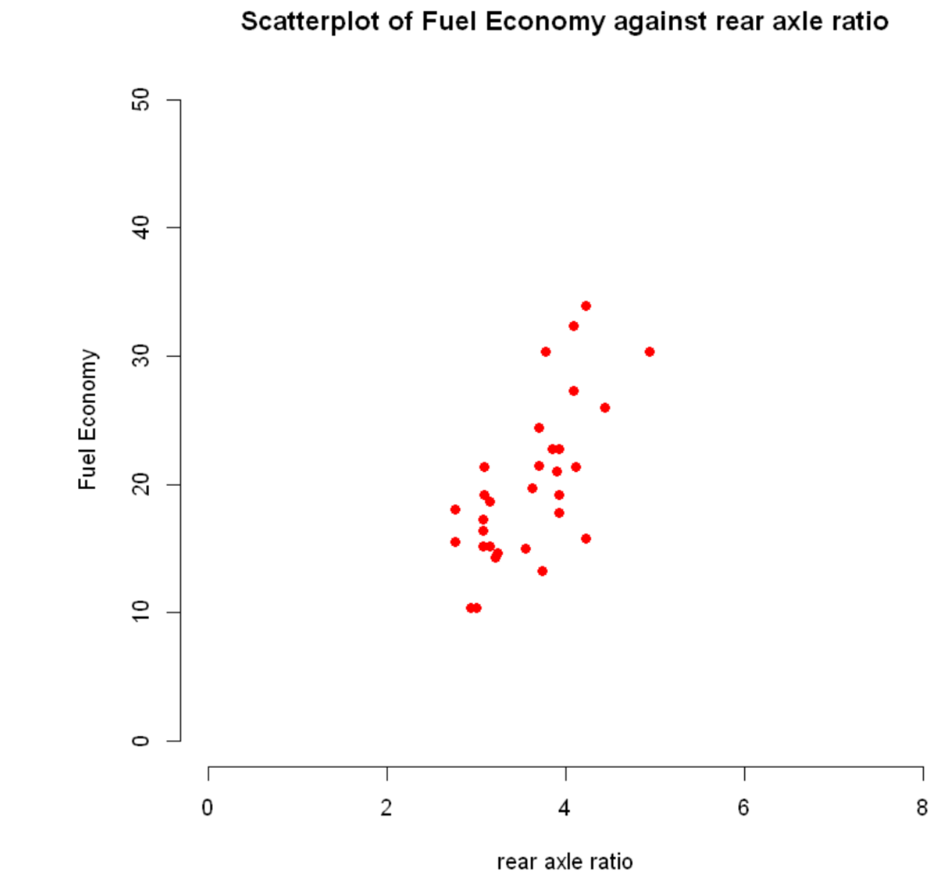
Gross horsepower measures the theoretical output of an engine's power output.

wt

Weight of the car per 1,000 lbs.

Correlation Analysis

Fuel Efficiency and Rear Axle Ratio:

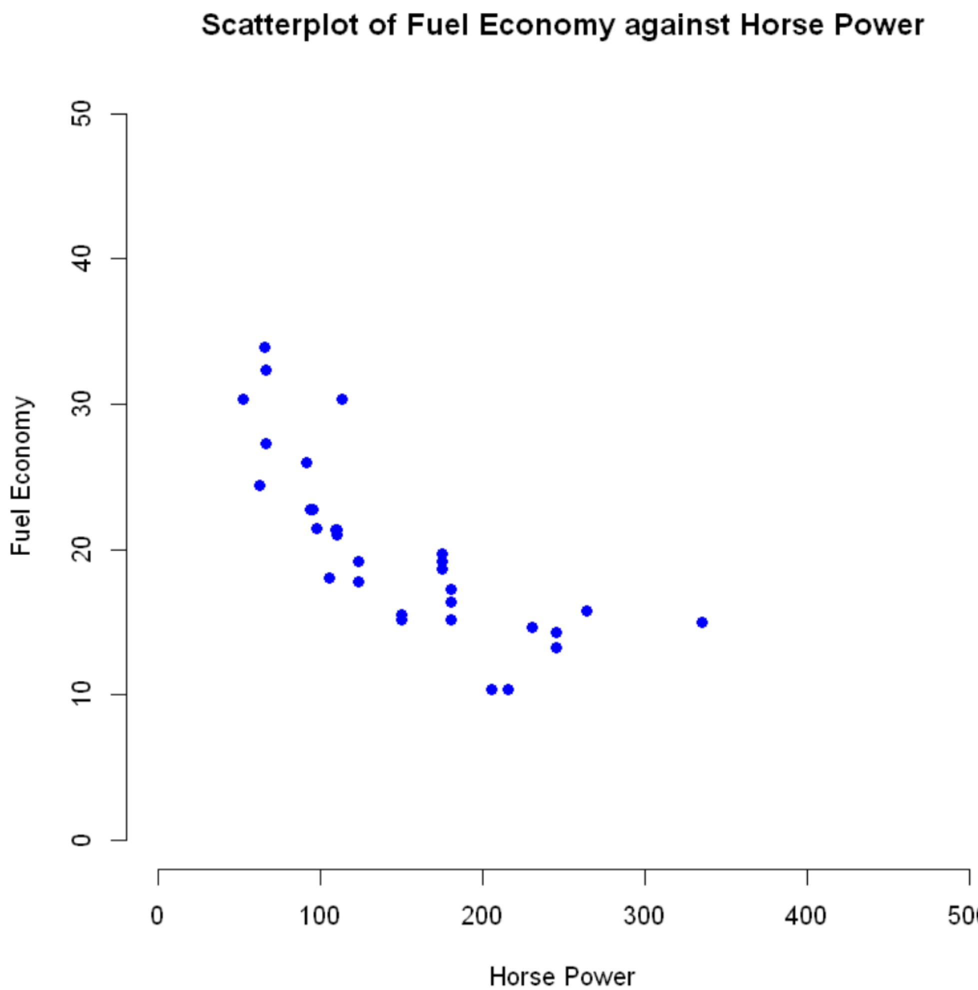


If the correlation coefficient between fuel efficiency and rear axle ratio is positive, this indicates a positive relationship between these two variables. A positive coefficient means that as the rear axle ratio increases, fuel efficiency also tends to increase. The strength of this relationship depends on how close the value is to 1. A value close to 0 indicates a weak relationship, whereas values approaching 1 signify a strong relationship.

drat and hp (-0.4488):

There is a moderate negative correlation between the rear axle ratio and horsepower. This suggests that cars with higher horsepower tend to have lower rear axle ratios. Lower ratios generally provide more torque and faster acceleration, which might be favored in high-performance vehicles that also typically feature higher horsepower.

Fuel Efficiency and Horsepower:



Similarly, a positive correlation coefficient between fuel efficiency and horsepower suggests that as the horsepower increases, fuel efficiency also increases. This is somewhat counterintuitive as traditionally; higher horsepower is associated with lower fuel efficiency due to greater power output and associated fuel consumption. However, in modern vehicles with advanced technologies like turbocharging and more efficient engines, higher horsepower does not necessarily result in lower fuel efficiency. Again, the strength of this correlation is determined by how close the coefficient value is to 1.

mpg and hp (-0.7762):

There is a strong negative correlation between mpg and horsepower. This indicates that as horsepower increases, fuel efficiency tends to decrease. This relationship is typical as higher horsepower often implies higher fuel consumption due to increased power output.

Results

```
Call:
lm(formula = mpg ~ drat + hp, data = mtcars_subset)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0369 -2.3487 -0.6034  1.1897  7.7500 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.789861   5.077752   2.125 0.042238 *  
drat         4.698158   1.191633   3.943 0.000467 ***
hp          -0.051787   0.009293  -5.573 5.17e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.17 on 29 degrees of freedom
Multiple R-squared:  0.7412,	Adjusted R-squared:  0.7233 
F-statistic: 41.52 on 2 and 29 DF,  p-value: 3.081e-09
```

R-squared and Adjusted R-squared

A multiple R-squared of 0.7412 - indicates that approximately 74.12% of the variability in fuel efficiency (mpg) can be explained by the linear relationship with rear axle ratio (drat) and horsepower (hp).

An adjusted R-squared of 0.7233 accounts for the number of predictors in the model (2 in this case) and indicates that about 72.33% of the variability in mpg is explained by the model, adjusting for the number of terms in the model. This is a good value suggesting that the model fits the data well without being overly fitted due to too many predictors.

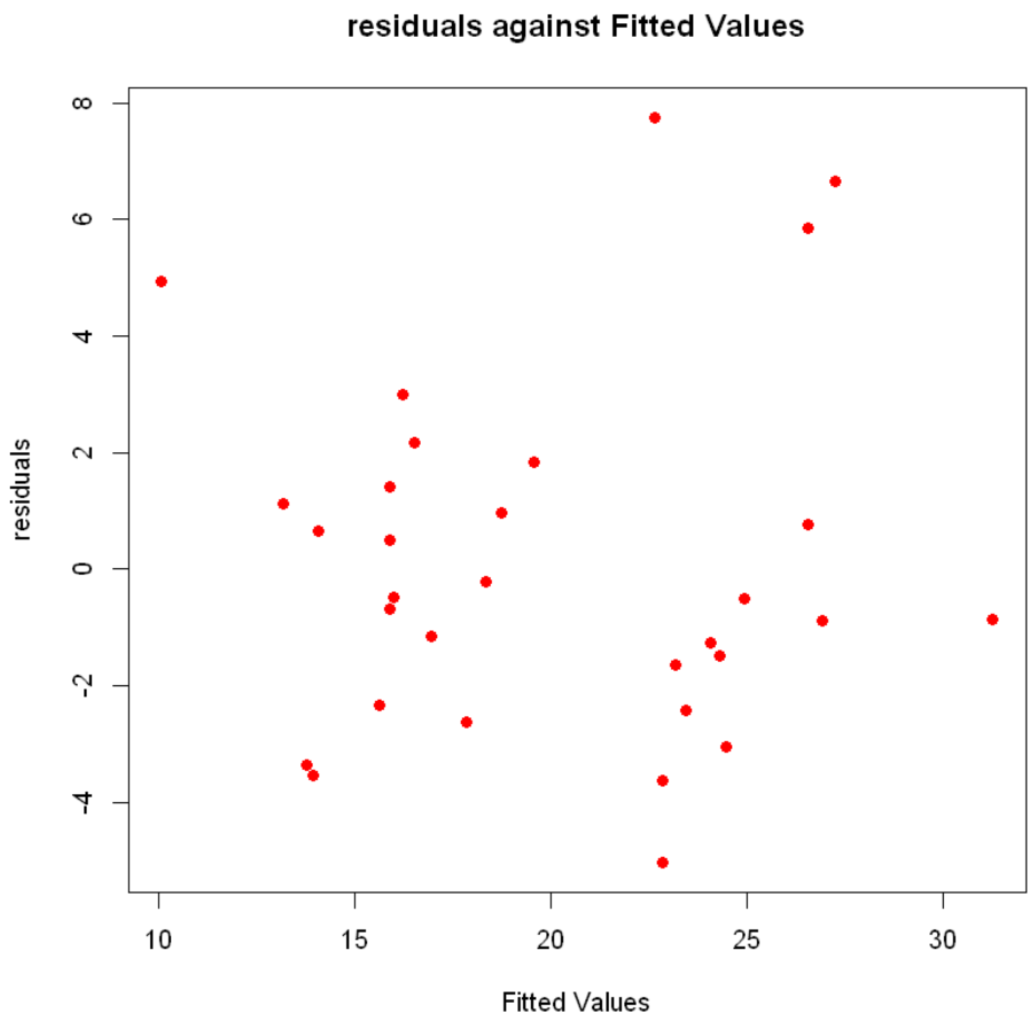
Beta Estimates

Intercept (10.789861): When both rear axle ratio and horsepower are zero (which is not practically possible), the predicted mpg is approximately 10.79.

The Rear Axle Ratio (drat, 4.698158): For each unit increase in the rear axle ratio, the fuel efficiency (mpg) is expected to increase by about 4.70 mpg, holding horsepower constant. This positive relationship is significant (p-value = 0.00467).

Horsepower (hp, -0.051787): For each unit increase in horsepower, the fuel efficiency (mpg) is expected to decrease by approximately 0.052 mpg, holding the rear axle ratio constant. This negative relationship is also highly significant (p-value = 5.17e-06).

Fitted and Residual Values

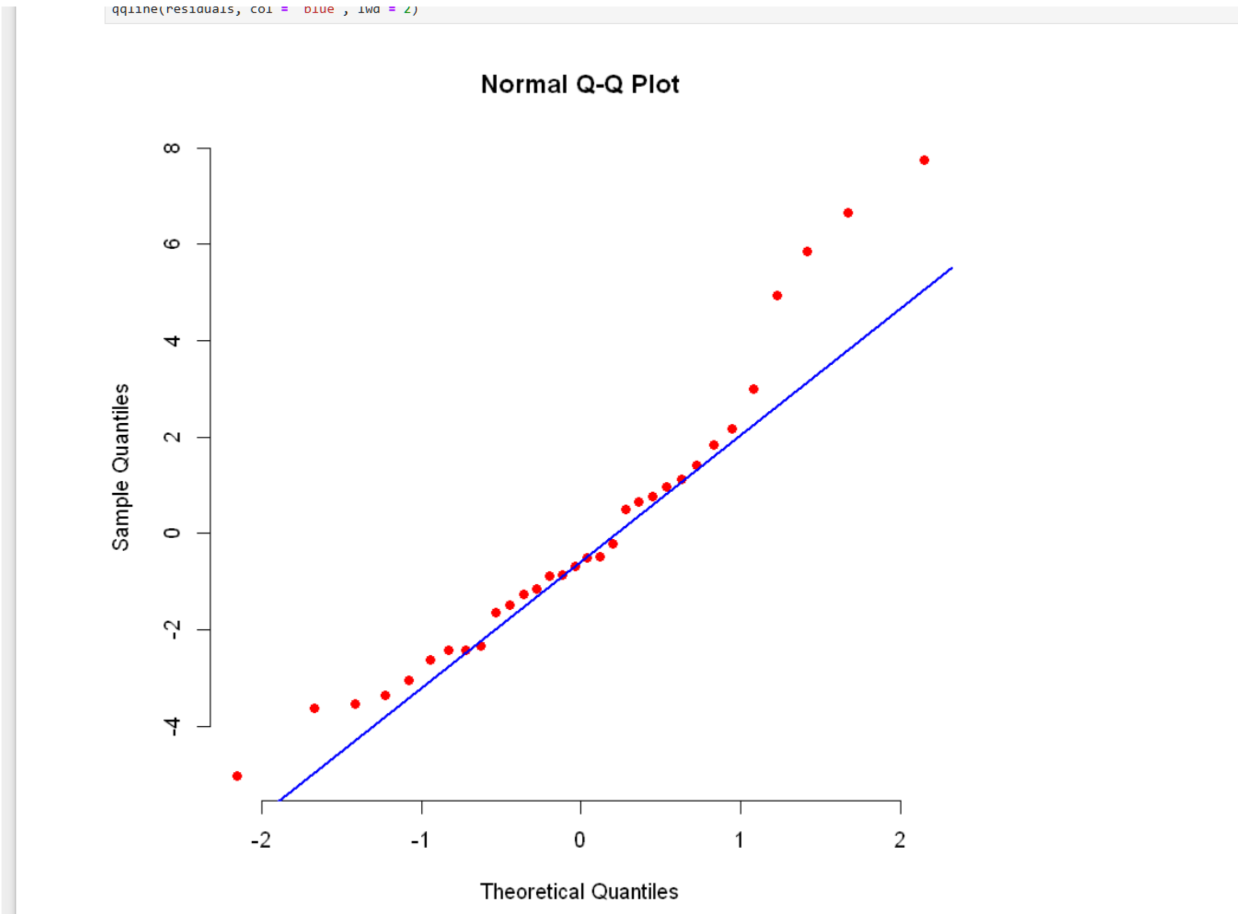


Fitted Value: This is the estimated value of the dependent variable (mpg) for each observation calculated from the regression equation using the known predictor values.

Residual: This is the difference between the observed value of mpg and the fitted value from the model. It represents the error of the estimate.

The "Residuals against Fitted Values" plot is critical for assessing the assumption of homoscedasticity in the regression model. Homoscedasticity means that the variances around the regression line are the same across all levels of the predictor variables. In a well-fitted model that meets this assumption. It is expected to see the residuals dispersed randomly around the horizontal line at zero without any noticeable patterns. Patterns such as funnels or curves suggest heteroscedasticity, meaning that the error variance changes at different levels of the predictors. This could lead to inefficient estimates and affect the validity of statistical tests.

Upon reviewing the plot provided and its random scatter points of residuals around the horizontal zero line the model does meet the homoscedasticity assumption. if there were visible patterns like a funnel shape, it would imply heteroscedasticity, requiring further investigation or adjustment to the model such as transforming variables.



The Normal Q-Q plot is used to check if the residuals from the regression model are normally distributed. In this plot, the quantiles of the residuals are plotted against the theoretical quantiles from a normal distribution. If the residuals are perfectly normal the points should lie exactly on the diagonal line. The points mostly follow the diagonal reference line, meaning that the residuals are approximately normally distributed. There are some deviations at the ends of the distribution This might indicate potential outliers or heavy tails in the distribution of residuals.

Model Significance

Overall Model Significance (F-test)

Null Hypothesis (H0): The model with no predictors (only the intercept) is sufficient to explain the relationship, all coefficients (except the intercept) are zero.

Alternative Hypothesis (H1): At least one predictor (either drat or hp) significantly affects the response variable mpg.

Test Statistics

F-statistic: 41.52

P-value: 3.081e-09

Since the p-value is much less than 0.05, we reject the null hypothesis at the 5% significance level. This indicates that the overall regression model is statistically significant, meaning that the predictors (rear axle ratio and horsepower) collectively provide a significantly better fit than a model with no predictors.

Rear Axle Ratio (drat)

Null Hypothesis (H0 for drat): Bdrat = 0 (drat does not affect mpg)

Alternative Hypothesis (H1 for drat): Bdrat ≠0 (drat affects mpg)

P-value for drat: 0.00467

Since the p-value for drat is less than 0.05, we reject the null hypothesis for drat. This indicates that the rear axle ratio is a significant predictor of mpg at the 5% significance level.

Horsepower (hp)

Null Hypothesis (H0 for hp):

βhp = 0 (hp does not affect mpg)

Alternative Hypothesis (H1 for hp):

βhp ≠ 0 (hp affects mpg)

P-value for hp: 5.17e-06

The p-value for hp is significantly less than 0.05, leading us to reject the null hypothesis for hp as well. This shows that horsepower significantly negatively impacts mpg.

95% Confidence Intervals for Parameter Estimates

A 95% confidence interval for **rear axle ratio** indicates that we are 95% confident the increase in mpg for each unit increase in drat lies within this range, demonstrating a significant positive impact on mpg.

A 95% confidence interval for **horsepower** indicates that we are 95% confident the decrease in mpg for each unit increase in horsepower lies within this range, demonstrating a significant negative impact.

Making Predictions Using the Model

Using the regression model to predict the fuel efficiency of a vehicle that has a rear axle ratio of 3.15 and a horsepower of 120 we use the coefficients obtained during the analysis and create the prediction equation $mpg=10.789861+4.698158\times drat-0.051787\times hp$ this equation illustrates a fuel efficiency of 19.38 miles per gallon. The residual is the difference between the observed value and the predicted value. Given that the observed mpg is 20.5 mpg (20.5mpg – 19.38mpg = 1.12mpg) This residual tells us that the actual mpg is 1.12 mpg higher than what the model predicts.

(Confidence Interval):

Fit: 19.37466 mpg

Lower Bound (lwr): 17.57162 mpg

Upper Bound (upr): 21.1777 mpg

(Prediction Interval):

Fit: 19.37466 mpg

Lower Bound (lwr): 12.64486 mpg

Upper Bound (upr): 26.10446 mpg

Interpretation

95% Confidence Interval for the Mean Response

Range: 17.57 to 21.18 mpg

We are 95% confident that the average fuel efficiency (mpg) for all cars with a rear axle ratio of 3.15 and horsepower of 120 lies between 17.57 and 21.18 mpg. This interval reflects the uncertainty about where the true mean mpg (for this combination of drat and hp) falls and is based on the variability of the sample data used to fit the model.

95% Prediction Interval for a Specific Observation

Range: 12.64 to 26.10 mpg

We are 95% confident that the actual fuel efficiency (mpg) of any individual car with a rear axle ratio of 3.15 and horsepower of 120 will fall between 12.64 and 26.10 mpg. This interval is wider because it accounts not only for the uncertainty in estimating the true mean mpg but also for the natural variability of mpg among cars with these specifications.: The prediction interval is wider than the confidence interval because it accounts for two types of uncertainties: the uncertainty in estimating the mean mpg (which the confidence interval also considers) and the additional variability in mpg among individual cars (which the confidence interval does not consider). The prediction interval must cover the variation around the predicted mean mpg due to individual differences among cars, making it necessarily broader. This wider range for the prediction interval provides a more practical and realistic estimate for what one might expect from an individual car's performance under specified conditions, reflecting both the accuracy of the model and the inherent variability of the outcome.

Conclusion

The regression analysis of the relationship between fuel efficiency (mpg) and two key variables—rear axle ratio (drat) and horsepower (hp)—has yielded several insights. The analysis has shown significant statistical results and provided valuable predictions for practical use. Based on the analysis and the results obtained, I would recommend using this model, assuming the sample size is sufficiently large and representative of the population for several reasons: The overall model is statistically significant as indicated by the F-test (p-value: 3.081e-09), suggesting that the model provides a good fit for the data compared to a model with no predictors.

Predictor Significance: Both predictors, drat and hp, are statistically significant with low p-values (0.00467 for drat and 5.17e-06 for hp), indicating that they both have a meaningful impact on mpg. The 95% confidence interval for the predicted mpg (17.57 to 21.18) for cars with a specific drat and hp suggests a high degree of certainty about the average effect of these variables on mpg. This interval helps in understanding the accuracy of the model's estimates and provides a range where the true mean likely falls. For automotive manufacturers, understanding how different specifications like drat and hp affect fuel efficiency helps in designing cars that balance performance with fuel economy leading to more marketable products and refined engineering.