Motor Trend MPG Data Analysis

Arnav Pant

1st November 2024

Executive Summary

This report explores the `mtcars` dataset to understand wether automatic or manual transmission vehicles give a better miles-per-gallon (mpg). A t-test shows manual transmission vehicles average 7.25 MPG more than automatic ones. However, after fitting a multiple linear regression model, manual transmission only adds about 1.81 MPG, with weight, horsepower, and cylinder count having a stronger effect on MPG.

Loading and Preparing Data

To start off, we will import the required data and convert certain columns to categorical (factor) data for easier analysis.

```r
library(ggplot2)
data(mtcars)
head(mtcars, n=3)
dim(mtcars)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
attach(mtcars)
```

Exploratory Analysis

The initial exploration of the mtcars dataset involves visualizing the relationship between transmission type and MPG. The goal is to see if there is an observable trend in MPG that varies based on whether the transmission is automatic or manual. The boxplot below provides a quick visual comparison of the MPG distributions for cars with automatic and manual transmissions. This plot serves as an initial, visual hypothesis: it allows us to see if there is a general trend in which one transmission type appears to be associated with a higher MPG. This plot, through Interpretation,allows us to conclude: If the boxplot shows a higher median MPG and a generally higher range of values for manual transmissions, this would give preliminary support to the idea that manual transmissions may yield better MPG on average

Figure I: Boxplot comparing MPG for automatic and manual transmissions. This plot suggests that manual transmissions generally achieve higher MPG. Gold for automatic, silver for manual

Statistical Test:

After observing the boxplot, we conduct a two-sample t-test to determine if the observed difference in MPG between automatic and manual transmissions is statistically significant. This test assesses whether the difference in MPG between these two transmission types is likely due to random variation or if it reflects a meaningful difference in fuel efficiency. T-Test for Transmission Type and MPG We perform a t-test to see if the MPG difference between transmission types is statistically significant.

```
testResults <- t.test(mpg ~ am)
testResults$p.value
```

```
## [1] 0.001373638
```

The small p-value from the t-test suggests a statistically significant MPG difference between the transmission types.

```
testResults$estimate
```

```
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

The t-test estimates that manual transmission vehicles have 7.25 more MPG than automatic ones.

Regression Analysis

Fitting a Full Model We begin the analysis by fitting a full linear regression model to explore the relationship between multiple car attributes (predictors) and miles per gallon (MPG). This comprehensive model considers all variables in the mtcars dataset to examine how each factor might affect MPG

```
fullModelFit <- lm(mpg ~ ., data = mtcars)
summary(fullModelFit)
```

we look for predictors with p-values below 0.05, as these indicate statistical significance. In this model, if no variables have a p-value below 0.05, it suggests that none of them have a clear, statistically significant impact on MPG, potentially due to multicollinearity or other factors. Since no p-values are below 0.05, we can't confirm which factors are strongly related to MPG.

Backward Selection for Significant Variables To create a more focused model, we use backward selection, a technique that iteratively removes the least significant predictors to retain only those with the strongest associations with MPG.

```
stepFit <- step(fullModelFit)
summary(stepFit)
```

Final Model Interpretation: The resulting model retains only the most impactful predictors: cylinders, horsepower, weight, and transmission type. With an R-squared of 0.87, this simplified model explains about 87% of the variation in MPG, indicating a strong fit.

Key results: - Adding 2 cylinders (from 4 to 6) reduces MPG by 3.03. - Adding 2 more cylinders (from 6 to 8) reduces MPG by 2.16. - Each 100 horsepower increase decreases MPG by 3.21. - Each 1000 lbs increase in weight decreases MPG by 2.5. - Manual transmission improves MPG by 1.81.

Model Diagnostics

Figure II: Model Diagnostic Plots

Diagnostic plots indicate: 1. Residuals vs. Fitted plot shows randomness, suggesting independence. 2. Normal Q-Q plot shows residuals are normally distributed. 3. Scale-Location plot shows consistent variance. 4. Residuals vs. Leverage plot indicates no significant outliers.

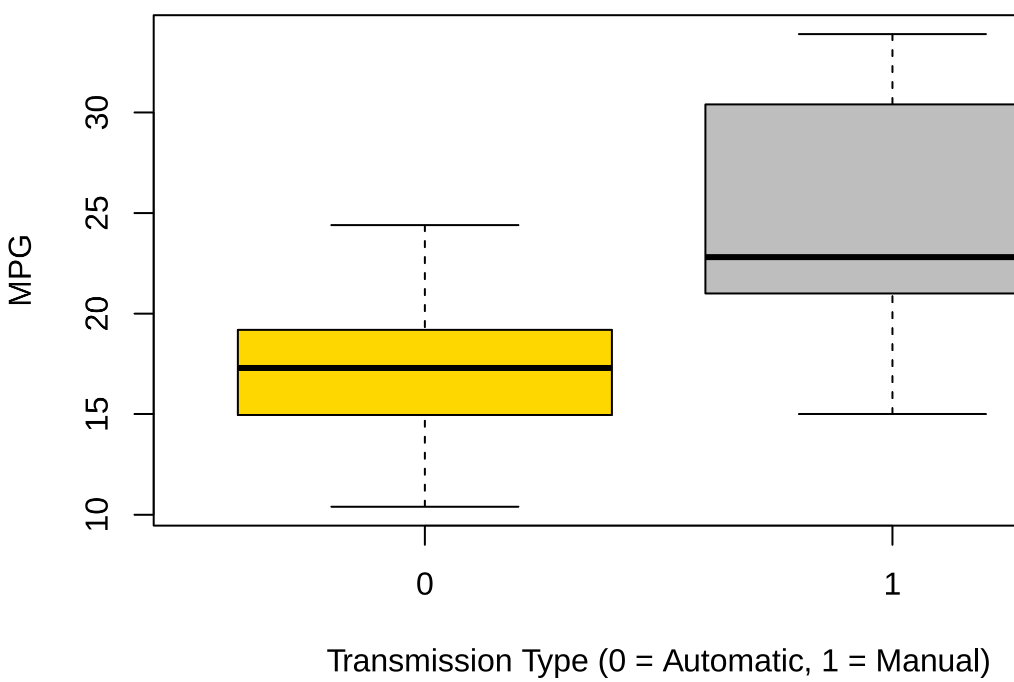
```r
sum((abs(dfbetas(stepFit))) > 1)
```

```
## [1] 0
```

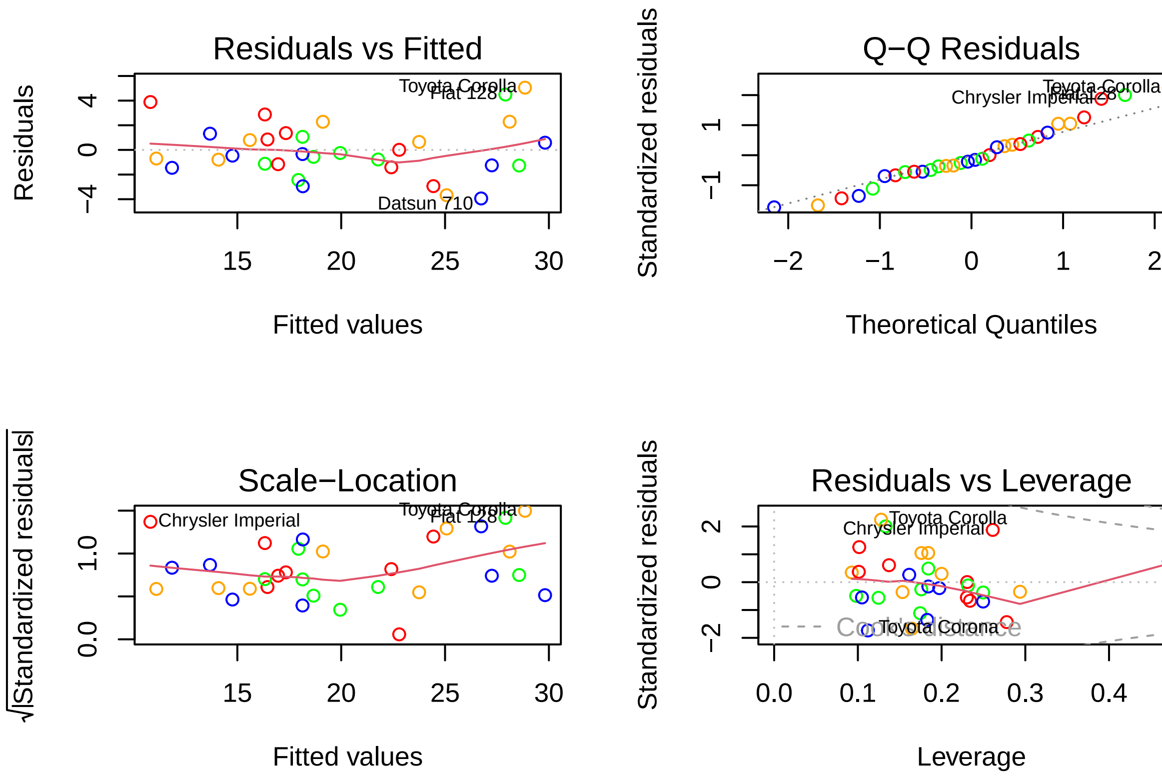
Conclusion

Manual transmissions offer a slight MPG advantage, but weight, horsepower, and cylinder count are more impactful in determining MPG.

Appendix Figures



I: MPG by Transmission Type



II: Diagnostic Plots