Motor Trends

Daniel Aklilu

2025-08-06

Executive Summary

This analysis explores whether manual or automatic transmissions are more fuel efficient using the mtcars dataset. Using the mtcars dataset, I began with the exploratory data analysis to compare MPG accross transmission types. I also used linear regression models, including both simple and multiple regression, to account for potential related variables such as weight, horsepower,and the numbers of cylinders. Finally I used Model diagnostics to perform the validity of the regression assumptions.

Introduction

-download mtcar data set and libraries

```
# Load Neccessary libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(dplyr)
library(ggplot2)
#install.packages("tinytext", repos = "https://cloud.r-projects.org/")
#tinytex::install_tinytex()
tinytex::tinytex_root()
#Load dataset
data("mtcars")
head(mtcars)
str(mtcars)
```

Exploratory Data analysis

```r
#Group by transmission and calculate summary statistics
convert <- mtcars %>%
      mutate(Transmission = factor(am, labels =
      c("Automatic", "Manual")))%>%
      group_by(Transmission) %>%
      summarise(Avg_mpg = mean(mpg),
      SD = sd(mpg),
      Total_cars = n())
# Display the summary table
head(convert)
```

Regression Modeling

```r
  # Create a new column for labeled
# transmission types
    mtcars_with_trans <- mtcars %>%
  mutate(Transmission = factor(am, labels
          = c("Automatic", "Manual")))


# Fit a simple linear regression model
 model1 <- lm(mpg ~ Transmission, mtcars_with_trans)
# Fit a multiple regression model controlling for weight, hp, and cyl.
 model2 <- lm(mpg ~ Transmission + wt + hp + cyl, mtcars_with_trans)

 model1
```

```
##
## Call:
## lm(formula = mpg ~ Transmission, data = mtcars_with_trans)
##
## Coefficients:
##        (Intercept)  TransmissionManual
##             17.147               7.245
```

```r
 model2
```

```
##
## Call:
## lm(formula = mpg ~ Transmission + wt + hp + cyl, data = mtcars_with_trans)
##
## Coefficients:
##        (Intercept)  TransmissionManual                  wt                  hp
##           36.14654             1.47805            -2.60648            -0.02495
##                cyl
##           -0.74516
```

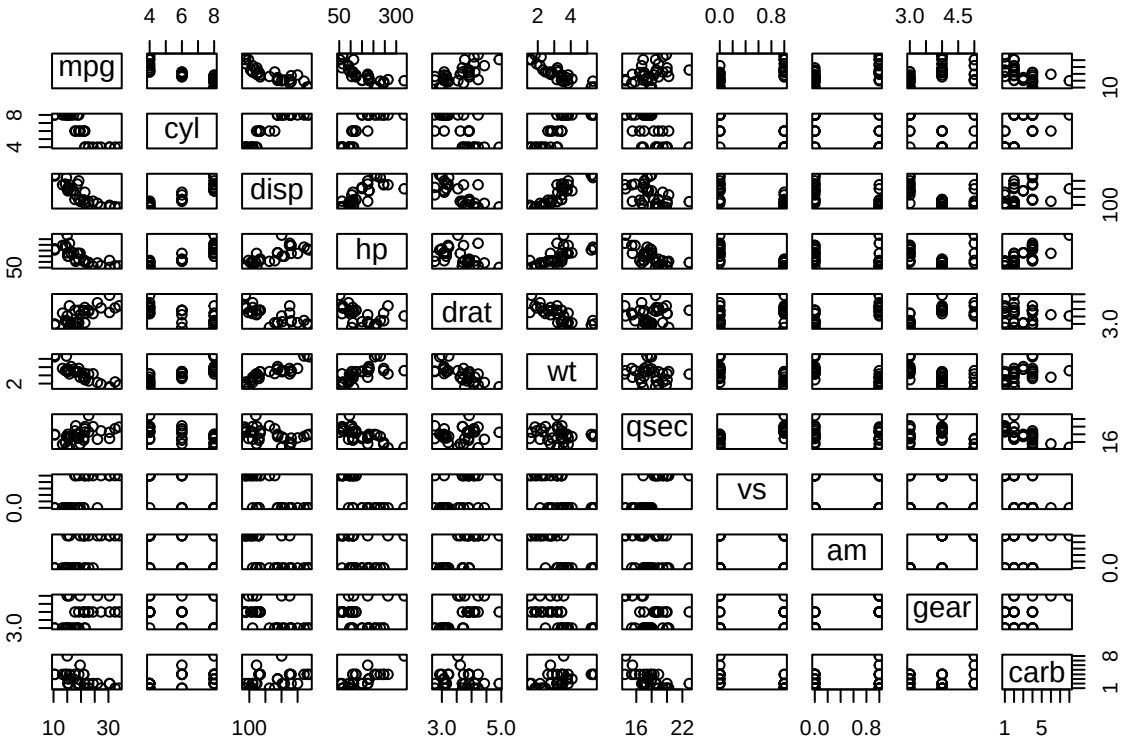
Linear Model

In the simple linear regression model, manual transmission cars had an estimated 7.25 MPG advantage over automatic cars. However, after adjusting for confounding factors such as weight, horsepower, and number of

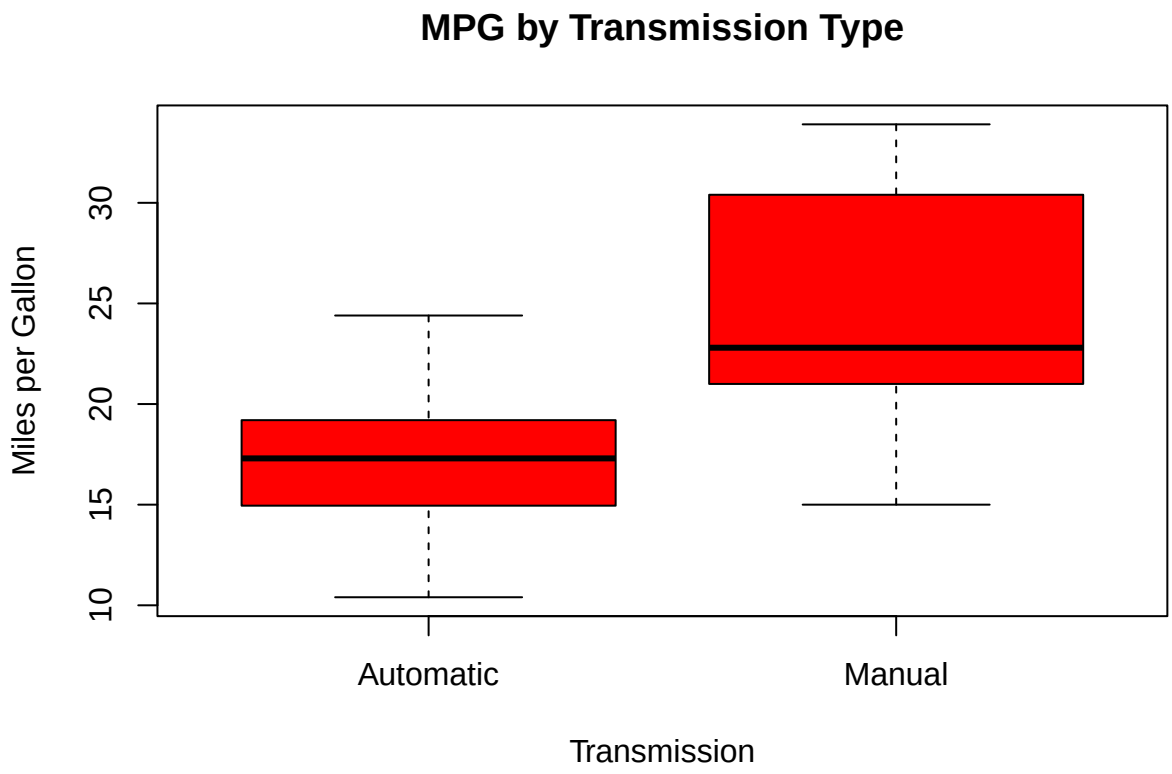
cylinders, the difference dropped to 1.48 MPG. This suggests that part of the initial difference is explained by other characteristics of the cars, though manual transmissions still appear to be more fuel efficient.

```
# Create a scatter plot matrix of all numeric variables
# in mtcars

# this helps me visually assess potential relationships
# and when variables give similar information.
pairs(mtcars)
```

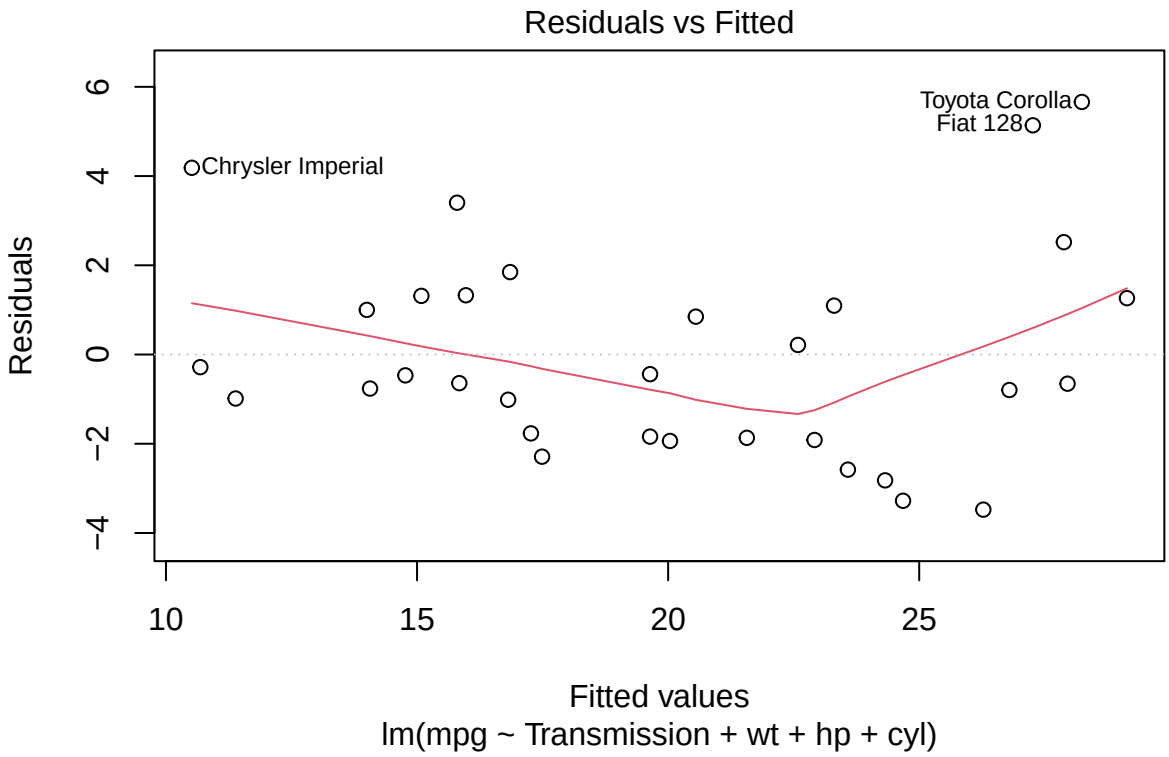


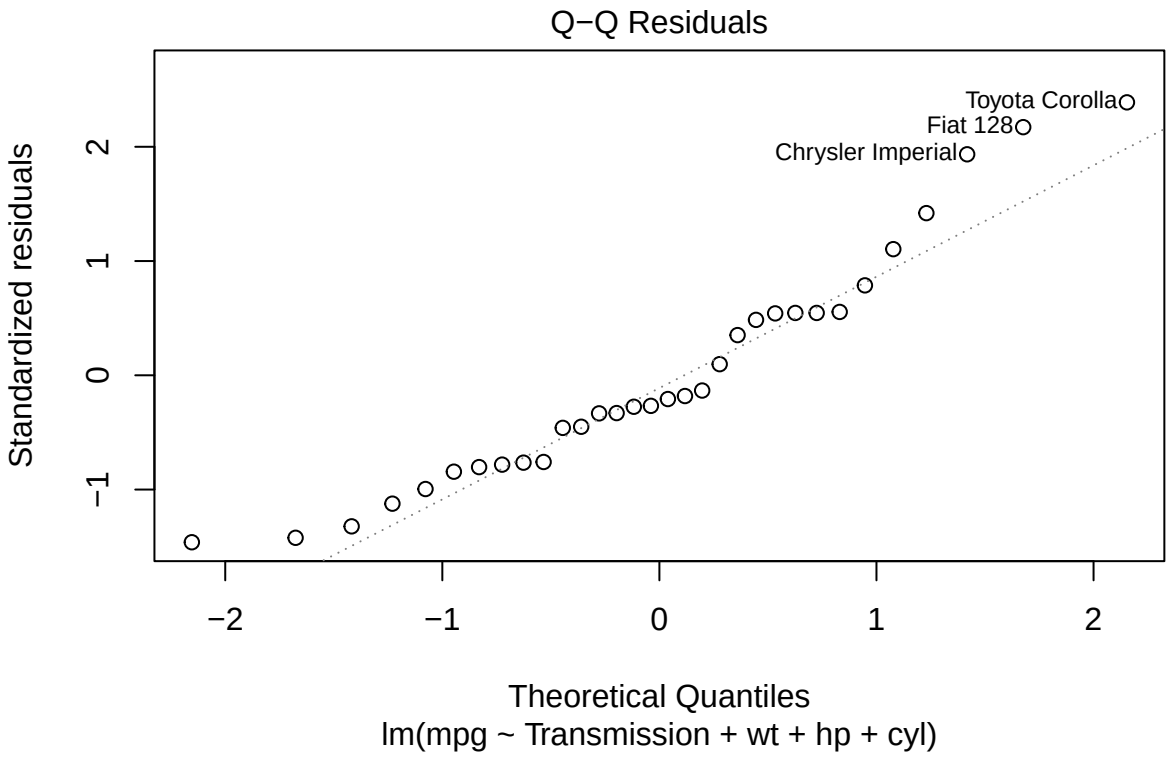

```
# Create a boxplot to visualize MPG by transmission type
boxplot(mpg ~ Transmission, mtcars_with_trans, xlab = "Transmission",
        ylab = "Miles per Gallon",
        main = "MPG by Transmission Type", col = "red")
```

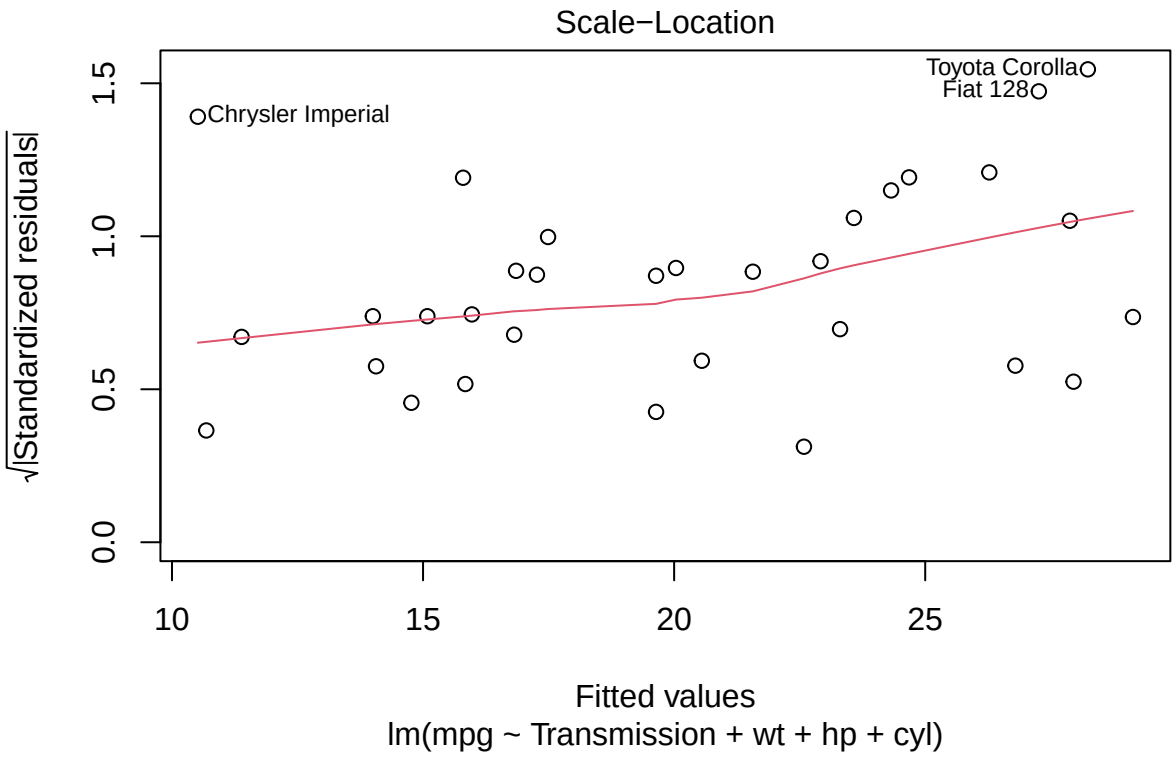


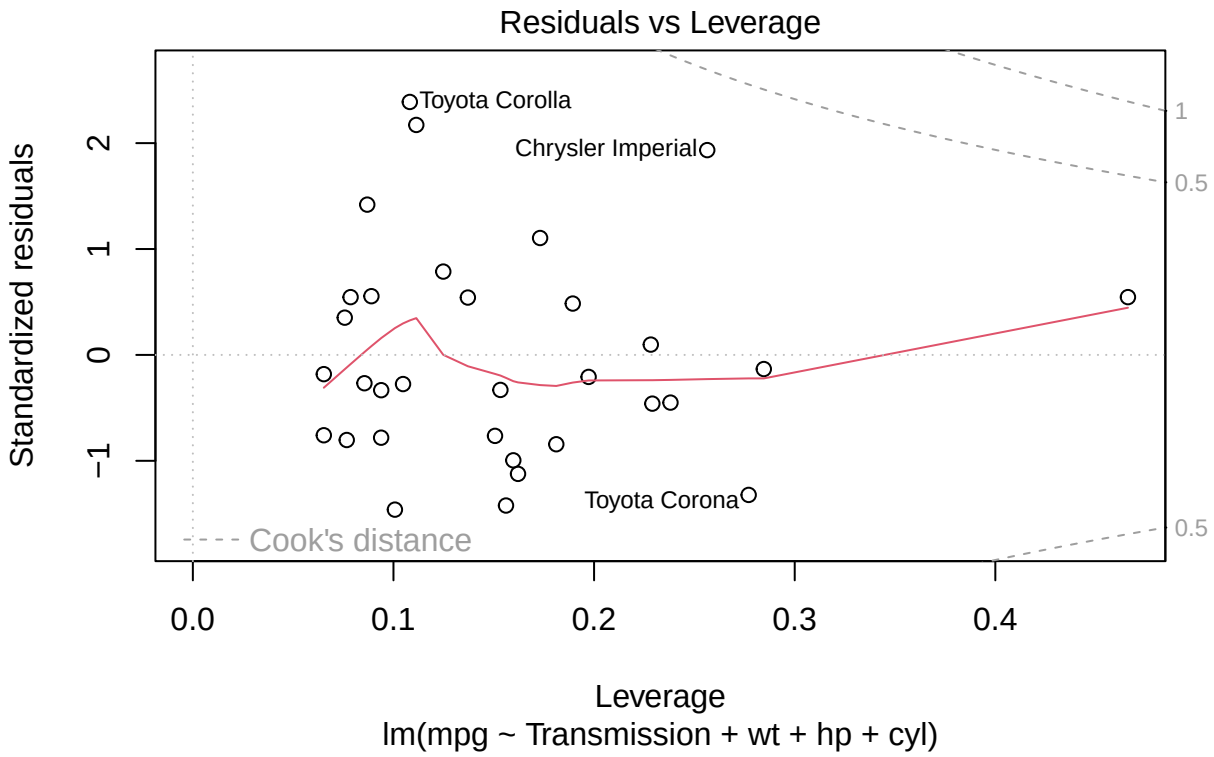
Residual Diagnostics

```
# Generate diagnostic plots to check model assumptions.
plot(model2)
```


Residuals vs Fitted
Toyota Corolla
Fiat 128
Chrysler Imperial
Residuals
6
4
2
0
−2
−4
10
15
20
25
Fitted values
lm(mpg ~ Transmission + wt + hp + cyl)


Q−Q Residuals
Toyota Corolla
Fiat 128
Chrysler Imperial
Standardized residuals
−1
0
1
2
−2
−1
0
1
2
Theoretical Quantiles
lm(mpg ~ Transmission + wt + hp + cyl)


Scale−Location
Toyota Corolla
Fiat 128
Chrysler Imperial
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Fitted values
lm(mpg ~ Transmission + wt + hp + cyl)



Summary I conducted residual diagnostics to evaluate whether the linear regression assumptions were satisfied. The residual vs fitted plot showed no major deviations from linearity. The Q-Q plot indicated that the residuals were normally distributed. The scale-location plot suggested roughly equal variance of residuals. Finally, no influential outliers were identified in the leverage plot. Finally, the model assumptions appear to be reasonably met.

```r
#show summary of model 2
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ Transmission + wt + hp + cyl, data = mtcars_with_trans)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4765 -1.8471 -0.5544  1.2758  5.6608
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        36.14654    3.10478  11.642 4.94e-12 ***
## TransmissionManual  1.47805    1.44115   1.026   0.3142
## wt                 -2.60648    0.91984  -2.834   0.0086 **
## hp                 -0.02495    0.01365  -1.828   0.0786 .
## cyl                -0.74516    0.58279  -1.279   0.2119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.509 on 27 degrees of freedom
## Multiple R-squared:  0.849,  Adjusted R-squared:  0.8267
## F-statistic: 37.96 on 4 and 27 DF,  p-value: 1.025e-10
```

```
# Show confidence interval for model
# coefficients
confint(model2)
```

```
##                           2.5 %       97.5 %
## (Intercept)        29.77605177 42.517019733
## TransmissionManual -1.47894635  4.435041763
## wt                 -4.49383134 -0.719130075
## hp                 -0.05295064  0.003048517
## cyl                -1.94093802  0.450623969
```

Statistical Inference results

A multiple linear regression was performed to estimat the effect of transmission type on MPG while controlling for weight, horsepower, and number of cylinders. The coefficient for manual transmission was 1.48 MPG, suggesting that, on average, manual cars had slightly higher fuel efficiency than automatic cars. However, the p-value associated with this estimate was 0.31, indicating that the difference was not statistically significant at the 5% level. The 95% confidence interval ranged from -1.44 to 4.40, which includes zero, providing further evidence that the observed difference may be due to chance. After I adjusted for other cars characteristics, there is insufficient evidence to conclude that transmission type has a significant impact on MPG.

Conclusion

Based on both exploratory analysis and regression modeling, we found that manual cars initially appeared to have better fuel efficiency. However, once we accounted for other car characteristics. For example, Weight, horsepower, and number of cylinders. The estimated difference decreased to 1.48 MPG and was not statistically significant. Our model diagnostics showed that linear regression assumptions were resonably satisfied. Overall, we conclude that transmission type alone doesn't significantly predict MPG in the presence of other influencing factors.