Regression Models Week 4 Project - Motor Trend MPG Analysis

Brandon Robinson

2024-05-12

Executive Summary

The report will analyze the relationship between transmission type (automatic / manual) and miles per gallon (MPG), and determine which type has more of an influence on the MPG of a car. The dataset used is `mtcars` which contains data of 32 cars between 1973 - 1974. The first step will be to conduct a t-test to determine if the null hypothesis is rejected. Next several linear regression models will be employed to determine the statistical significance of variables against the MPG of cars.

Load Data

The required libraries will be loaded.

```
library(ggplot2)
library(GGally)
```

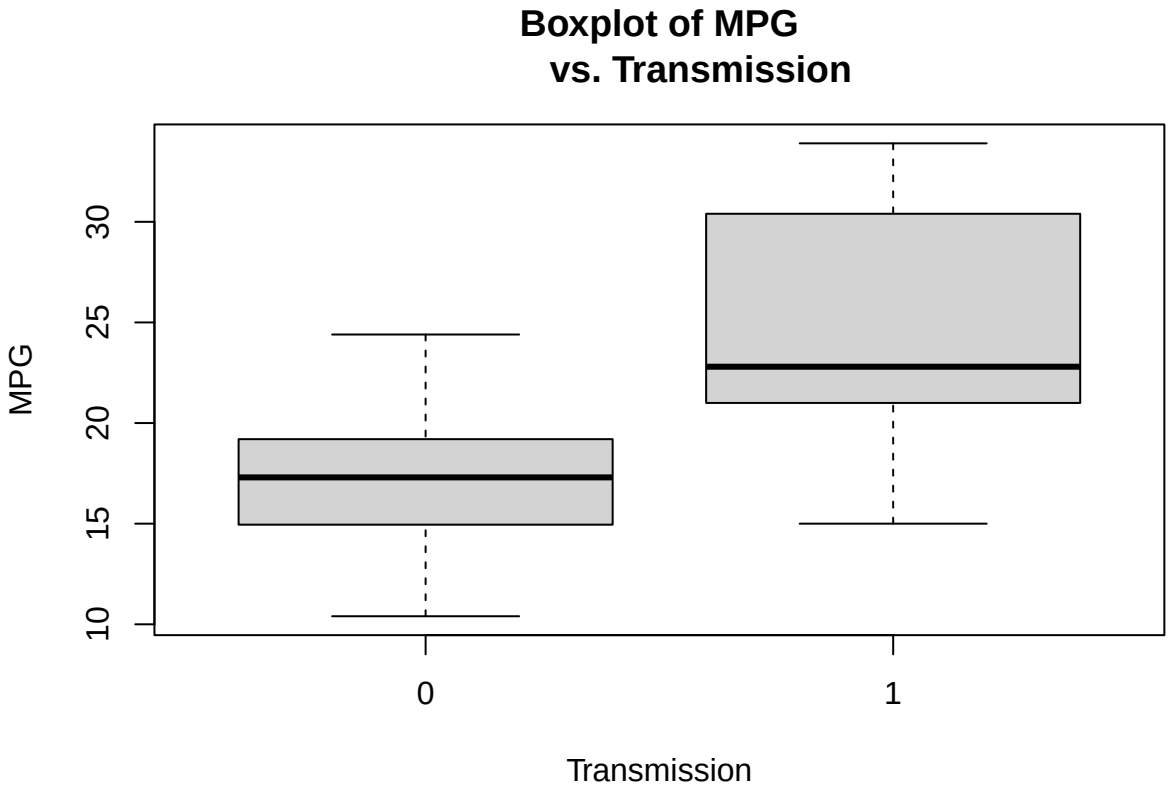
```
data(mtcars)
head(mtcars, n = 6) # Display the first 6 rows of the mtcars dataset
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Exploratory analysis

The next step is to perform initial exploratory analysis of the cars dataset and change some factors from `numeric` to `factor` class.

```
boxplot(mtcars$mpg ~ mtcars$am, xlab = "Transmission", ylab = "MPG", main = "Boxplot of MPG
        vs. Transmission")
```



From the intial boxplot one can infer manual transmission has a greater influence on the MPG

After the initial analysis, several variables will be changed in order to explore the influence on MPG.

```
mtcars$am <- gsub("0", "auto", mtcars$am)
mtcars$am <- gsub("1", "manual", mtcars$am)
mtcars$am <- factor(mtcars$am)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Statistical Inference

A t-test will be performed to verify if the null hypothesis should be rejected

```
tResults <- t.test(mpg ~ am, mtcars)
tResults$p.value
```

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

```
tResults$estimate
```

```
##   mean in group auto mean in group manual
##             17.14737             24.39231
```

The p-value is 0.00137, so the null hypothesis is rejected. In addition, the estimate shows the MPG of manual cars is ~7 more than automatic.

Regression Analysis

The first step is to look at the impact of all variables on the full dataset.

```
full_model <- lm(mpg ~ ., mtcars)
summary(full_model)$r.squared
```

```
## [1] 0.8930749
```

The full model shows a 89% variance of mpg on all variables, but this does not provide us clear detail on which variables have the largest impact.

Next, only two variables were analysed, automatic and manual, to see if this has the largest influence on the MPG

```
am_model <- lm(mpg ~ am, mtcars)
summary(am_model)$r.squared
```

```
## [1] 0.3597989
```

Even though the transmission type affects MPG, the variance is only 36% which means there are other variables that have more of an influence

The next test is to use the step function to determine the variables which have the greatest influence over MPG.

```
step_model <- step(full_model, direction = "backward", trace = FALSE)
summary(step_model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## ammanual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

The step function included the additional variables, cyl6, cyl8, hp, and wt alongside transmission.

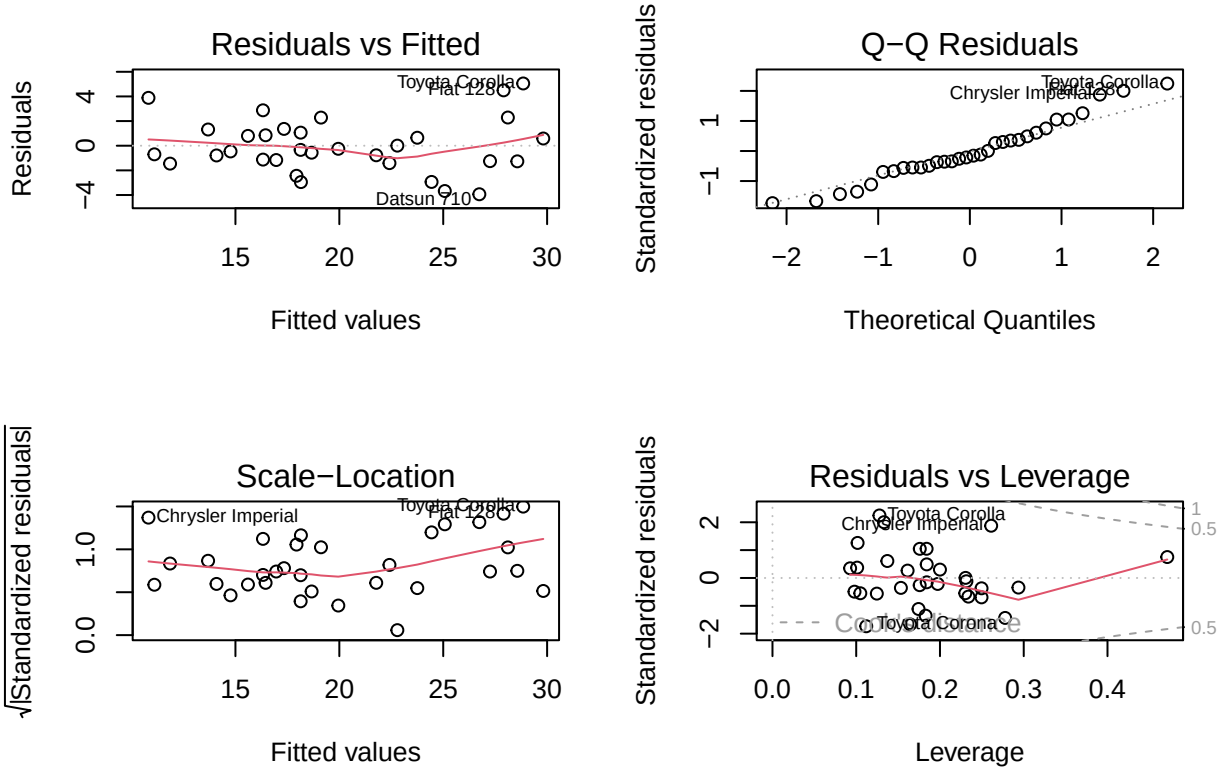
Conclusion

Based on the results from `step_model` it can be inferred other variables have more of an influence on the cars MPG. If the additional variables, cyl6, cyl8, hp, and wt, are held constant the manual transmissions get 1.8 MPG more on average than automatic.

Appendix

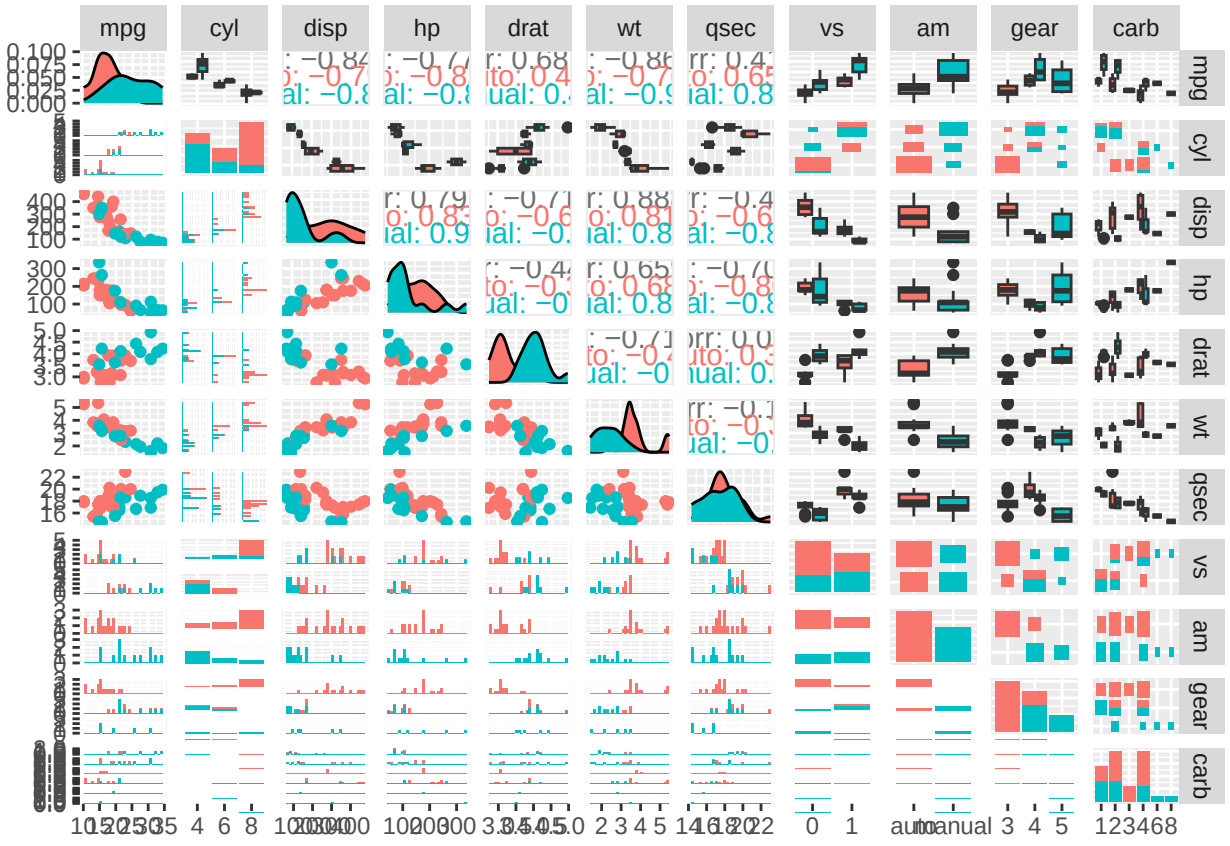
1. Residual plots

```
par(mfrow = c(2,2))
plot(step_model)
```



2. Pair graph

```
ggpairs(mtcars, mapping = aes(colour = am))
```


mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
manual