Motor Trend Analysis

Danish Tamboli

8/22/2020

Summary

In this Analysis we are going to look at the mtcars dataset (Motor Trend Car Road Tests), The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

The Questions that are of interest are:

- Is an automatic or manual transmission better for MPG?
- What is the MPG difference between automatic and manual transmissions?

Loading the Dataset:

```
data <- mtcars
```

Quick glance at Dataset:

```
head(data)
```

```
##                    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
```

```
str(data)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
```

```
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Upon looking at the structure of the Dataset, We notice that there are certain Variables that would be more useful if converted into Factors, They are:

- Number of Cylinders (cyl)
- Engine V-shaped (0) vs straight (1) (vs)
- Transmission Type Automatic (0) vs Manual (1) (am)
- Number of forward gears (gear)
- Number of Carburetors (carb)

Refer to Fig-2 in Appendix.

Conversion into Factors:

```
data$cyl <- as.factor(data$cyl)
data$vs <- as.factor(data$vs)
data$am <- factor(data$am)
data$gear <- as.factor(data$gear)
data$carb <- as.factor(data$carb)
```

Initals Inference:

```
temp_dat <- data[,c("mpg","am")]
temp_dat$am <- factor(temp_dat$am,labels = c("Automatic","Manual"))
head(temp_dat)
```

```
##                    mpg        am
## Mazda RX4         21.0    Manual
## Mazda RX4 Wag     21.0    Manual
## Datsun 710        22.8    Manual
## Hornet 4 Drive    21.4 Automatic
## Hornet Sportabout 18.7 Automatic
## Valiant           18.1 Automatic
```

```
knitr::kable(aggregate(mpg~am,data = temp_dat,mean),
             col.names = c("Transmission Type","Average Miles per Gallon (MPG)"))
```

Transmission Type	Average Miles per Gallon (MPG)
Automatic	17.14737
Manual	24.39231

```
difference <- aggregate(mpg~am,data = temp_dat,mean)[2,2] - aggregate(mpg~am,data = temp_dat,mean)[1,2]
```

The Difference in MPG between Manual and Automatic Transmission is **7.2449393**, This shows that Manual Cars perform more efficiently than Automatic Cars. This will be our Hypothesis against the NUll Hypothesis that there is no difference in MPG between the two Transmission types.

Refer to Fig-1 in Appendix.

```
dat_aut <- data[data$am=="0",]
dat_ml <- data[data$am=="1",]
print(test <- t.test(dat_aut$mpg,dat_ml$mpg))
```

```
##
##  Welch Two Sample t-test
##
## data:  dat_aut$mpg and dat_ml$mpg
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

The P-value is **0.0013736**, Indicating that the two lie in difference Distributions and hence there is a significant difference in MPG between the two Transmission Types.

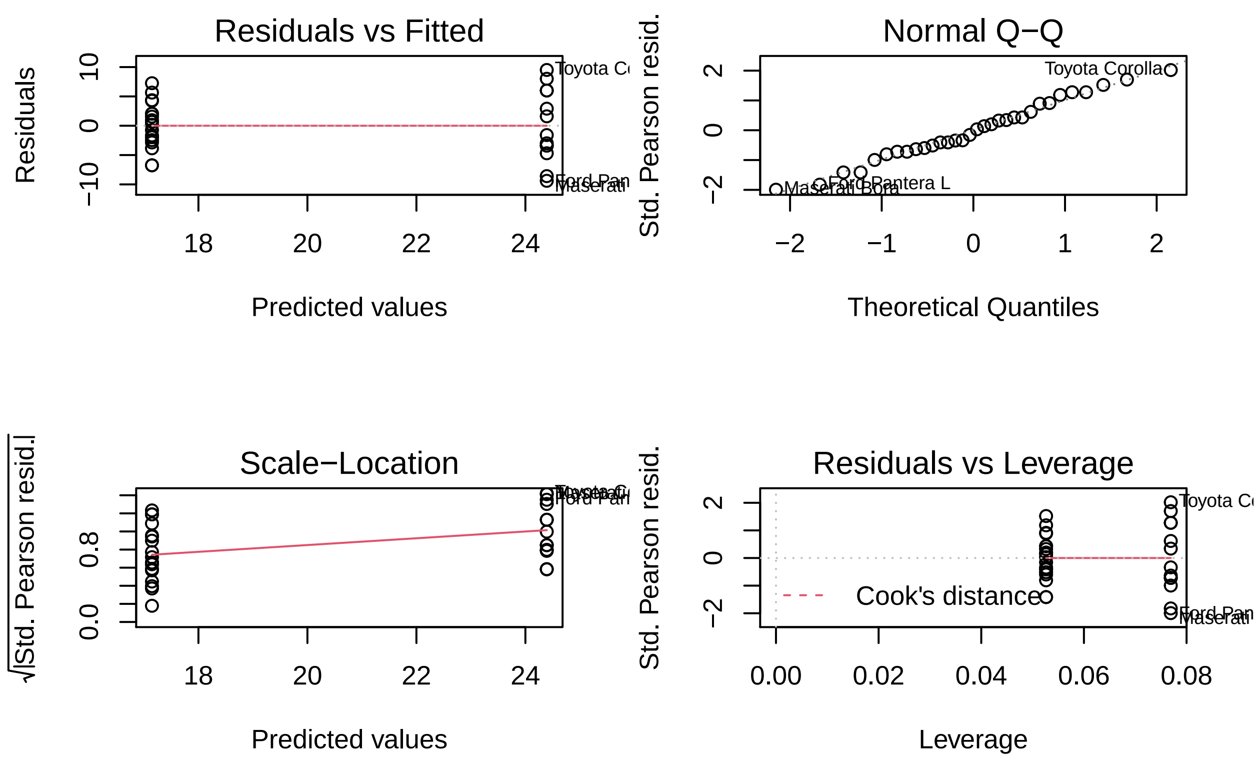
Regression Models

```
fit1 <- glm(mpg~am-1,family = "gaussian",data = data)
summary(fit1)
```

```
##
## Call:
## glm(formula = mpg ~ am - 1, family = "gaussian", data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -9.3923  -3.0923  -0.2974   3.2439   9.5077
##
## Coefficients:
##     Estimate Std. Error t value Pr(>|t|)
## am0   17.147      1.125   15.25 1.13e-15 ***
## am1   24.392      1.360   17.94  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 24.02989)
##
##     Null deviance: 14042.3  on 32  degrees of freedom
```

```
## Residual deviance:  720.9  on 30  degrees of freedom
## AIC: 196.48
##
## Number of Fisher Scoring iterations: 2
```

```r
par(mfrow=c(2,2))
plot(fit1)
```



The First Regression Model shows a similar story, with the difference in MPG between Manual and Transmission being **7.2449393**.

```r
summary(aov(mpg~.,data = data))
```

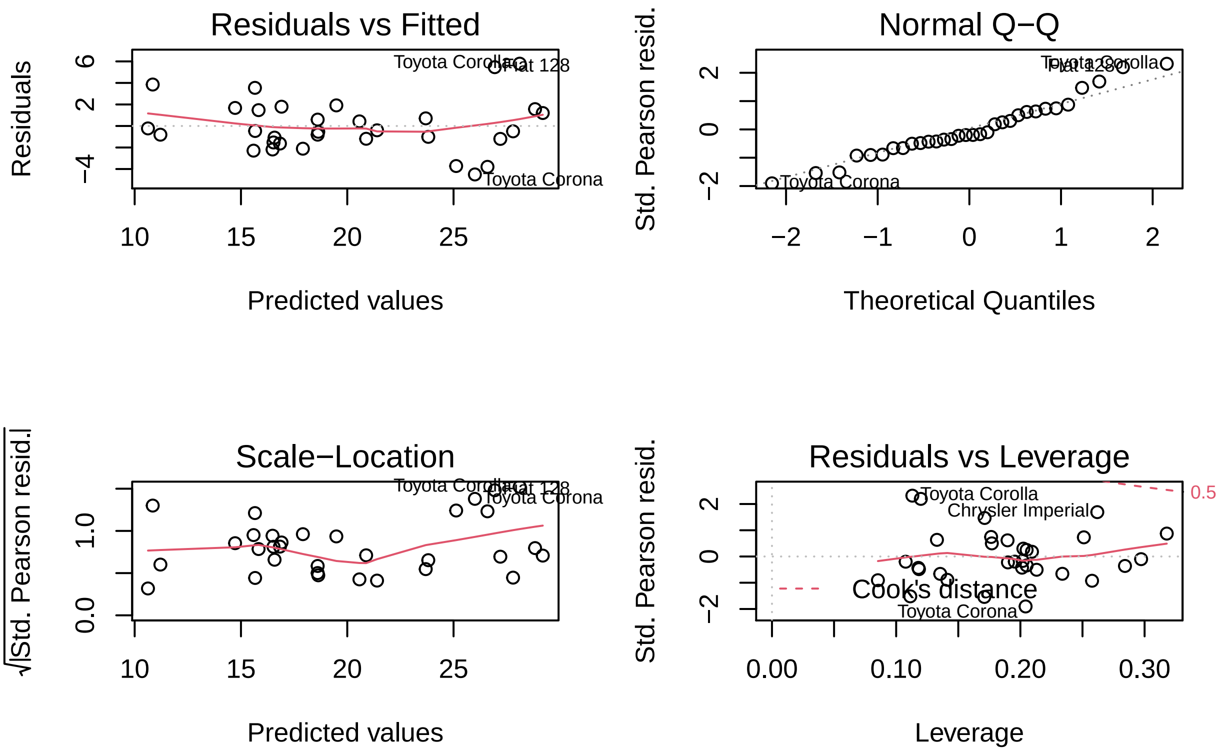
```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          2  824.8   412.4  51.377 1.94e-07 ***
## disp         1   57.6    57.6   7.181   0.0171 *
## hp           1   18.5    18.5   2.305   0.1497
## drat         1   11.9    11.9   1.484   0.2419
## wt           1   55.8    55.8   6.950   0.0187 *
## qsec         1    1.5     1.5   0.190   0.6692
## vs           1    0.3     0.3   0.038   0.8488
## am           1   16.6    16.6   2.064   0.1714
## gear         2    5.0     2.5   0.313   0.7361
## carb         5   13.6     2.7   0.339   0.8814
## Residuals   15  120.4     8.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis of Variance shows that we could have a better explained Model if we include other factors such as: * Number of Cylinder (cyl) * Displacement (disp) * Weight (1000lbs) (wt)

```
fit2 <- glm(mpg~am+cyl+disp+wt-1, family="gaussian",data=data)
summary(fit2)
```

```
##
## Call:
## glm(formula = mpg ~ am + cyl + disp + wt - 1, family = "gaussian",
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.5029  -1.2829  -0.4825   1.4954   5.7889
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## am0  33.816067   2.914272  11.604 8.79e-12 ***
## am1  33.957279   2.151443  15.783 7.78e-15 ***
## cyl6 -4.304782   1.492355  -2.885  0.00777 **
## cyl8 -6.318406   2.647658  -2.386  0.02458 *
## disp  0.001632   0.013757   0.119  0.90647
## wt   -3.249176   1.249098  -2.601  0.01513 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 7.033436)
##
##     Null deviance: 14042.31  on 32  degrees of freedom
## Residual deviance:   182.87  on 26  degrees of freedom
## AIC: 160.59
##
## Number of Fisher Scoring iterations: 2
```

```
par(mfrow=c(2,2))
plot(fit2)
```



The Second model is indeed a better explained model. With this we can conclude that Yes, There is indeed a distinction in MPG with respect to Transmission type, but also with respect to:

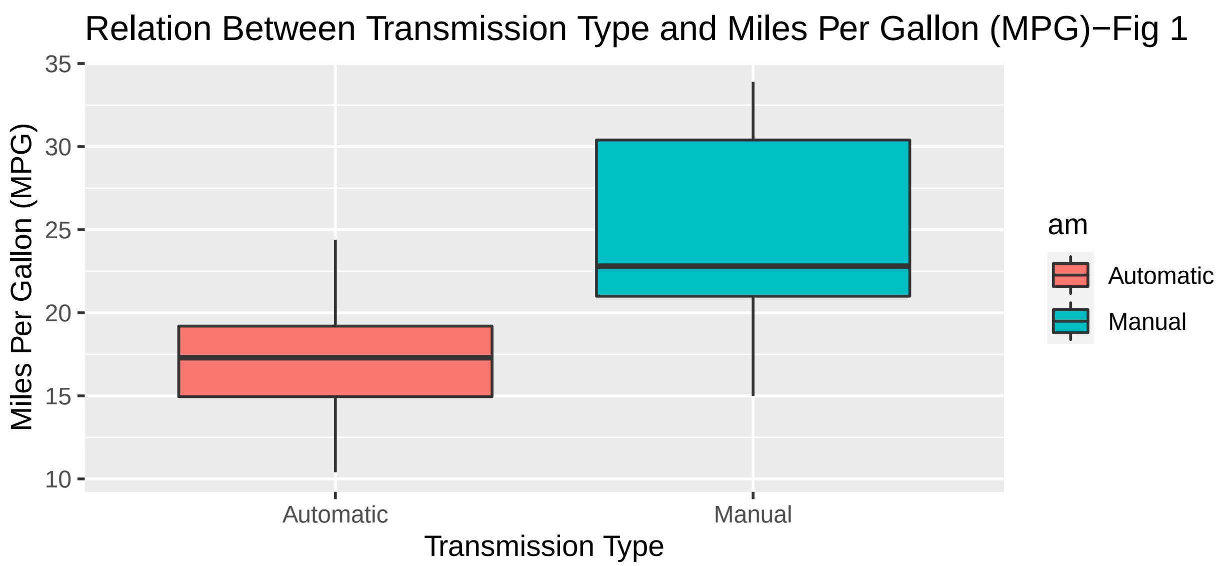
- Number of Cylinders (cyl)
- Displacement of the Engine (disp)
- The Total Weight of the Vehicle (wt)

Appendix

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
ggplot(data = temp_dat) + geom_boxplot(aes(am,mpg,fill=am)) + labs(x="Transmission Type",y = "Miles Per
```




```r
pairs(mpg~.,data = data,main="Scatterplot Matrix -Fig 2")
```