Peer Graded Assignment for Motor Trend

Dipali Bagad

September 3, 2017

Executive Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions: 1. Is an automatic or manual transmission better for MPG 2. Quantify the MPG difference between automatic and manual transmissions

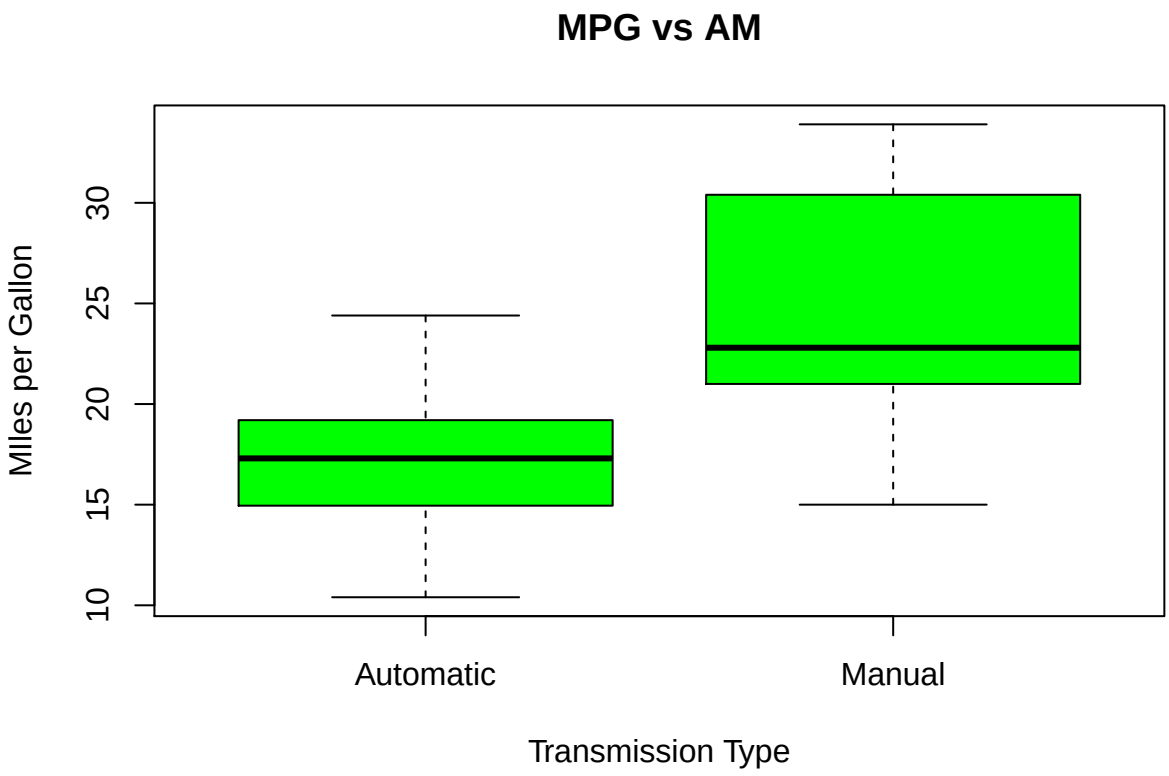
Data Processing

For analysis, first we need to load the dataset

```
library(datasets)
data(mtcars)
# transform variables to factor format
mtcars$am[mtcars$am==0]<-"Automatic"
mtcars$am[mtcars$am==1]<-"Manual"
mtcars$am<-as.factor(mtcars$am)
mtcars$cyl<-as.factor(mtcars$cyl)
mtcars$vs<-as.factor(mtcars$vs)
mtcars$gear<-as.factor(mtcars$gear)
mtcars$carb<-as.factor(mtcars$carb)
#read few records
head(mtcars)
```

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

```
#draw boxplot for transmission type comparision
boxplot(mtcars$mpg~as.factor(mtcars$am),col="green",ylab="MIles per Gallon",xlab="Transmission Type",ma
```



From graph we can say that manual transmissions have better mpg ratio than automatic transmissions. we can test this using welch test

```
t.test(mpg~am,mtcars,paired=FALSE,var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## 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 in group Automatic    mean in group Manual
##                17.14737                24.39231
```

The p-value is less than 0.05, so this confirms the hypothesis. we need to set basic and full model with possible variable and check which model influences more using step function and backward direction

```
basic<-lm(mpg~am,mtcars)
full<-lm(mpg~.,mtcars)
fit<-step(full, direction="backward")
```

```
## Start:  AIC=76.4
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
```

```
## - carb  5   13.5989 134.00 69.828
## - gear  2    3.9729 124.38 73.442
## - am    1    1.1420 121.55 74.705
## - qsec  1    1.2413 121.64 74.732
## - drat  1    1.8208 122.22 74.884
## - cyl   2   10.9314 131.33 75.184
## - vs    1    3.6299 124.03 75.354
## <none>              120.40 76.403
## - disp  1    9.9672 130.37 76.948
## - wt    1   25.5541 145.96 80.562
## - hp    1   25.6715 146.07 80.588
##
## Step:  AIC=69.83
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##        Df Sum of Sq    RSS    AIC
## - gear  2    5.0215 139.02 67.005
## - disp  1    0.9934 135.00 68.064
## - drat  1    1.1854 135.19 68.110
## - vs    1    3.6763 137.68 68.694
## - cyl   2   12.5642 146.57 68.696
## - qsec  1    5.2634 139.26 69.061
## <none>              134.00 69.828
## - am    1   11.9255 145.93 70.556
## - wt    1   19.7963 153.80 72.237
## - hp    1   22.7935 156.79 72.855
##
## Step:  AIC=67
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - drat  1    0.9672 139.99 65.227
## - cyl   2   10.4247 149.45 65.319
## - disp  1    1.5483 140.57 65.359
## - vs    1    2.1829 141.21 65.503
## - qsec  1    3.6324 142.66 65.830
## <none>              139.02 67.005
## - am    1   16.5665 155.59 68.608
## - hp    1   18.1768 157.20 68.937
## - wt    1   31.1896 170.21 71.482
##
## Step:  AIC=65.23
## mpg ~ cyl + disp + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - disp  1    1.2474 141.24 63.511
## - vs    1    2.3403 142.33 63.757
## - cyl   2   12.3267 152.32 63.927
## - qsec  1    3.1000 143.09 63.928
## <none>              139.99 65.227
## - hp    1   17.7382 157.73 67.044
## - am    1   19.4660 159.46 67.393
## - wt    1   30.7151 170.71 69.574
##
```

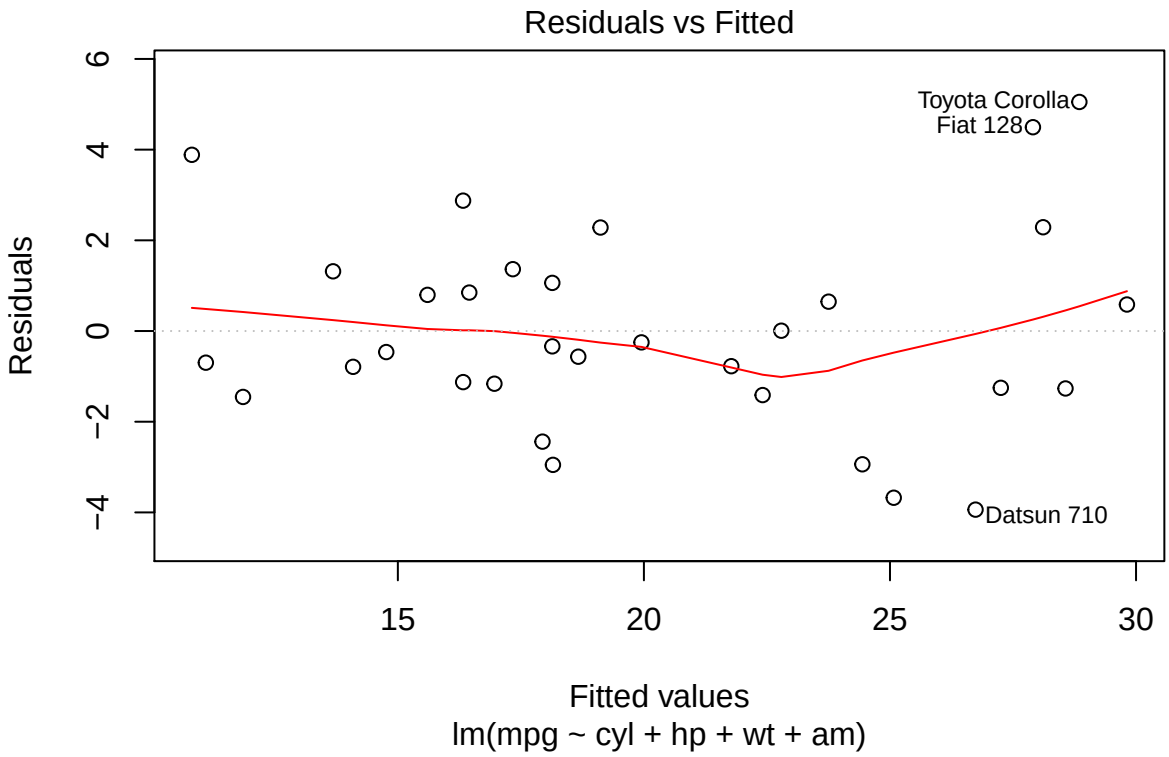
```
## Step:  AIC=63.51
## mpg ~ cyl + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - qsec  1     2.442 143.68 62.059
## - vs    1     2.744 143.98 62.126
## - cyl   2    18.580 159.82 63.466
## <none>              141.24 63.511
## - hp    1    18.184 159.42 65.386
## - am    1    18.885 160.12 65.527
## - wt    1    39.645 180.88 69.428
##
## Step:  AIC=62.06
## mpg ~ cyl + hp + wt + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - vs    1     7.346 151.03 61.655
## <none>              143.68 62.059
## - cyl   2    25.284 168.96 63.246
## - am    1    16.443 160.12 63.527
## - hp    1    36.344 180.02 67.275
## - wt    1    41.088 184.77 68.108
##
## Step:  AIC=61.65
## mpg ~ cyl + hp + wt + am
##
##        Df Sum of Sq    RSS    AIC
## <none>              151.03 61.655
## - am    1     9.752 160.78 61.657
## - cyl   2    29.265 180.29 63.323
## - hp    1    31.943 182.97 65.794
## - wt    1    46.173 197.20 68.191
```

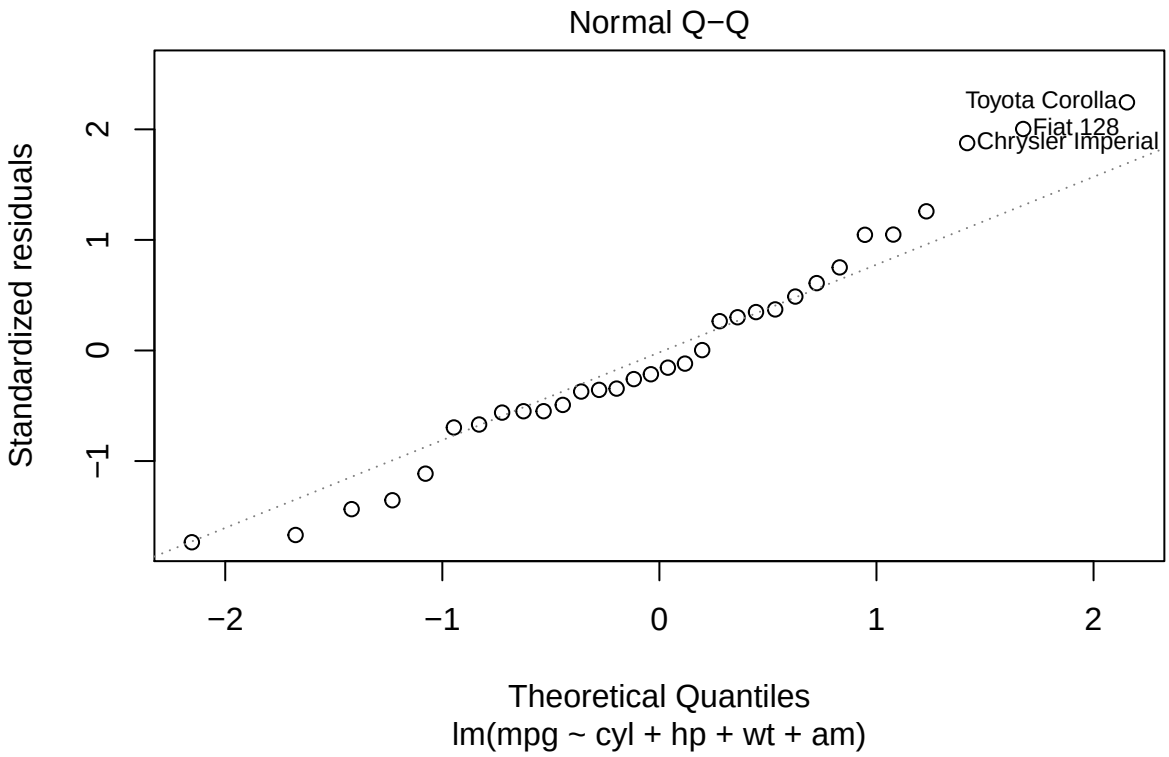
By Anova comparison we can see that very low f value. , we can say that its good to use the fit model instead of basic. The adjusted R squared is also much better (0.84 fit vs 0.34 basic)

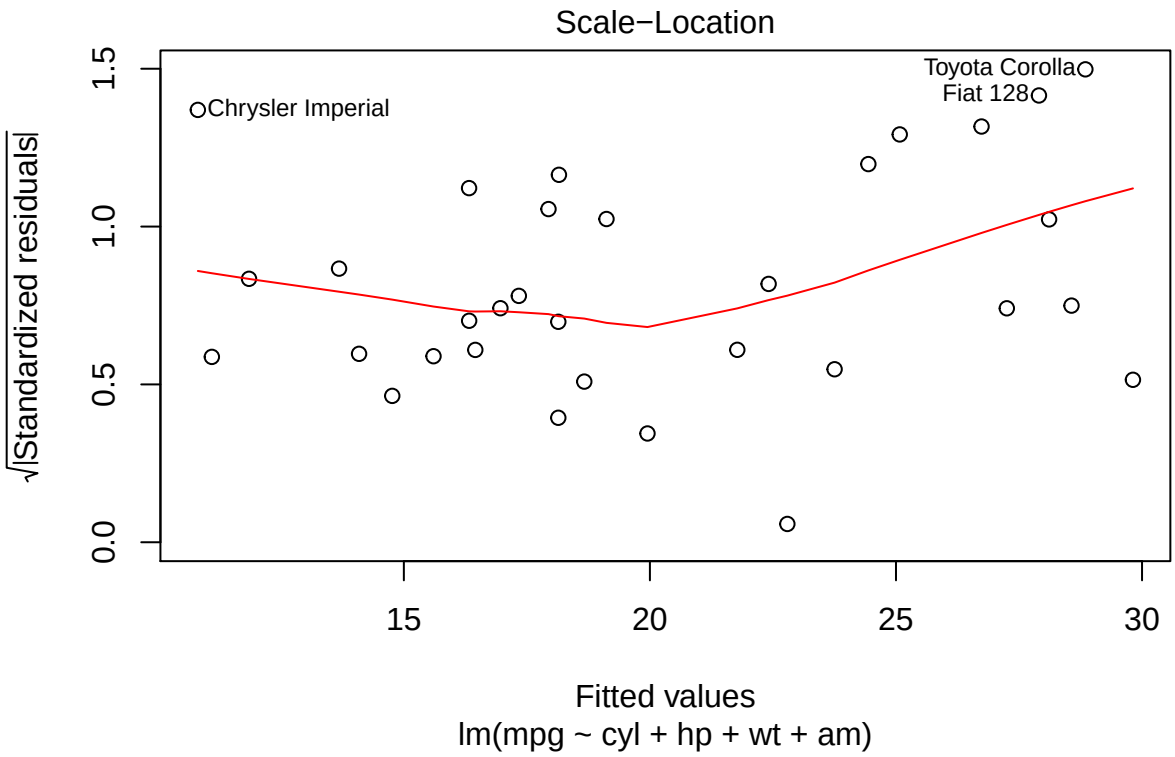
```
anova(basic,fit)
```

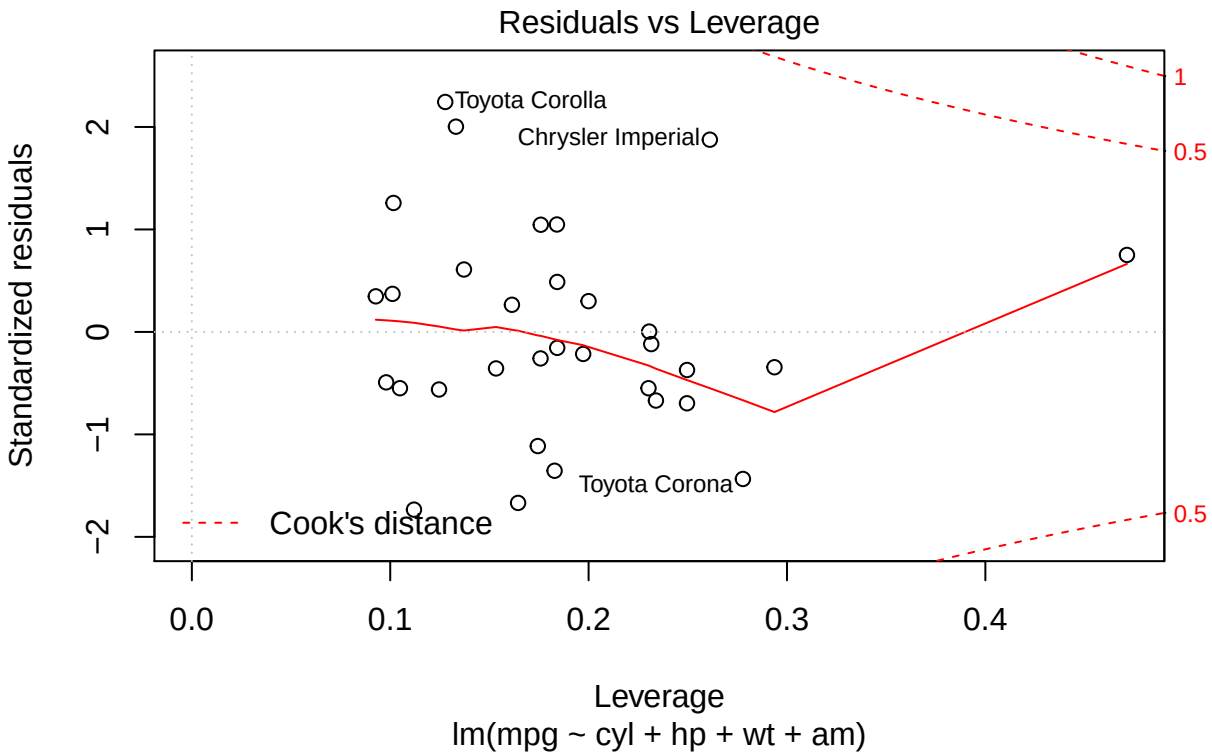
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(1, 1))
plot(fit)
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ cyl + hp + wt + am)
Toyota Corolla
Fiat 128
Datsun 710
6
4
2
0
−2
−4
15
20
25
30


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


Scale−Location
Chrysler Imperial
Toyota Corolla
Fiat 128
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
30
Fitted values
lm(mpg ~ cyl + hp + wt + am)



Residuals vs Fitted and Scale-Location plots show no pattern, the Normal Q-Q plot indicates that Residuals approximately follow a Normal distributions, and the Residuals vs Leverage plot tells that there's none particular outlier to be concerned.

Conclusion

Based on analysis, we can conclude that manual transmissions are better than automatic by 1.80921 mpg. This is the diffence between the intercept of the fit model and Manual.