Relationship between a set of variables and MPG

Dongjun_Cho

7/7/2020

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:

- Is an automatic or manual transmission better for MPG
- Quantify the MPG difference between automatic and manual transmissions

Analysis

Brief analysis of dataset

```
data <- mtcars
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
```

```
dim(data)
```

```
## [1] 32 11
```

```
summary(data)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
```

```
##   Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##   1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##   Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##   Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##   3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##   Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am             gear            carb
##   Min.   :0.0000   Min.   :3.000   Min.   :1.000
##   1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##   Median :0.0000   Median :4.000   Median :2.000
##   Mean   :0.4062   Mean   :3.688   Mean   :2.812
##   3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##   Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

Multivariable Regression / Interpretation

```r
summary(lm(mpg ~. -1, data))$coef
```

```
##           Estimate Std. Error      t value   Pr(>|t|)
## cyl    0.35082641 0.76292423   0.45984438 0.65014009
## disp   0.01354278 0.01762273   0.76848373 0.45037109
## hp    -0.02054767 0.02143989  -0.95838513 0.34828334
## drat   1.24158213 1.46276742   0.84878985 0.40513967
## wt    -3.82613150 1.86238084  -2.05443023 0.05200271
## qsec   1.19139689 0.45942323   2.59324480 0.01659185
## vs     0.18972068 2.06824861   0.09173011 0.92774262
## am     2.83222230 1.97512820   1.43394353 0.16564985
## gear   1.05426253 1.34668717   0.78285629 0.44205756
## carb  -0.26321386 0.81235653  -0.32401273 0.74898869
```

Data Format

- mpg : Miles/(US) gallon
- cyl : Number of cylinders
- disp : Displacement (cu.in.)
- hp : Gross horsepower
- drat : Rear axle ratio
- wt : Weight (1000 lbs)
- qsec : 1/4 mile time
- vs : Engine (0 = V-shaped, 1 = straight)
- am : Transmission (0 = automatic, 1 = manual)
- gear : Number of forward gears
- carb : Number of carburetors

```r
pairs(data, panel=panel.smooth, main ="Car")
```

Car
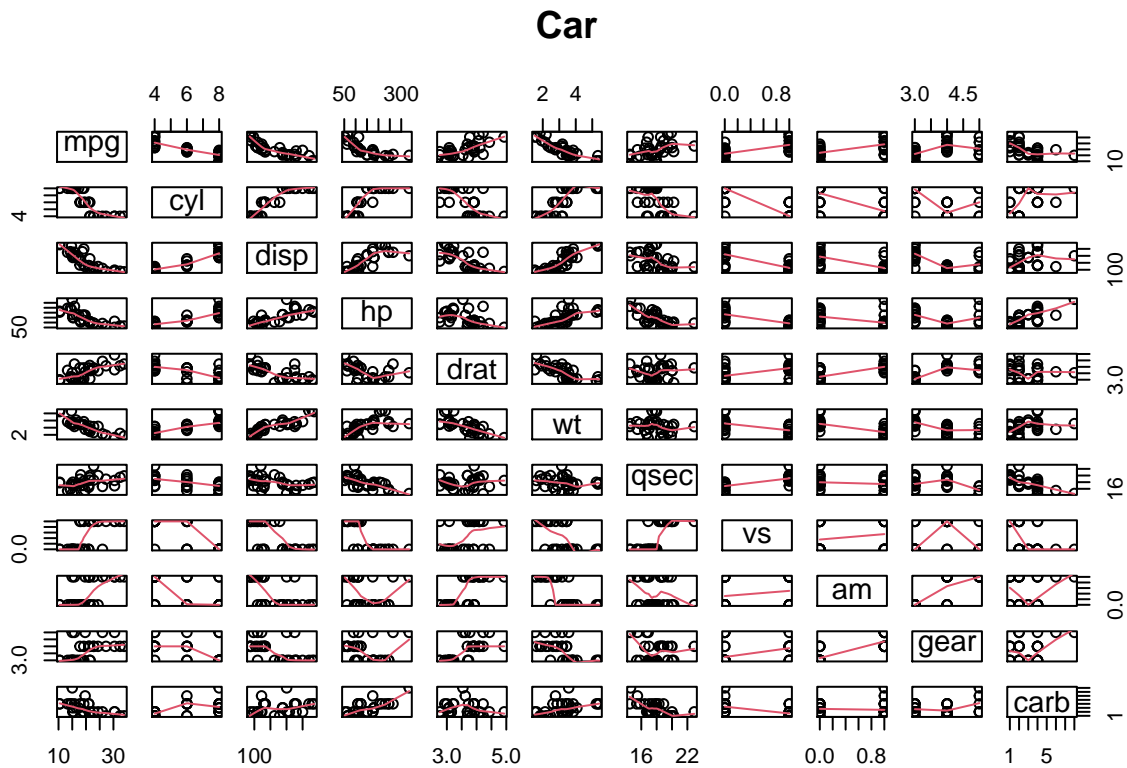
MPG by type of transmission

Automatic

```
summary(data[data$am == 0, "mpg"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```
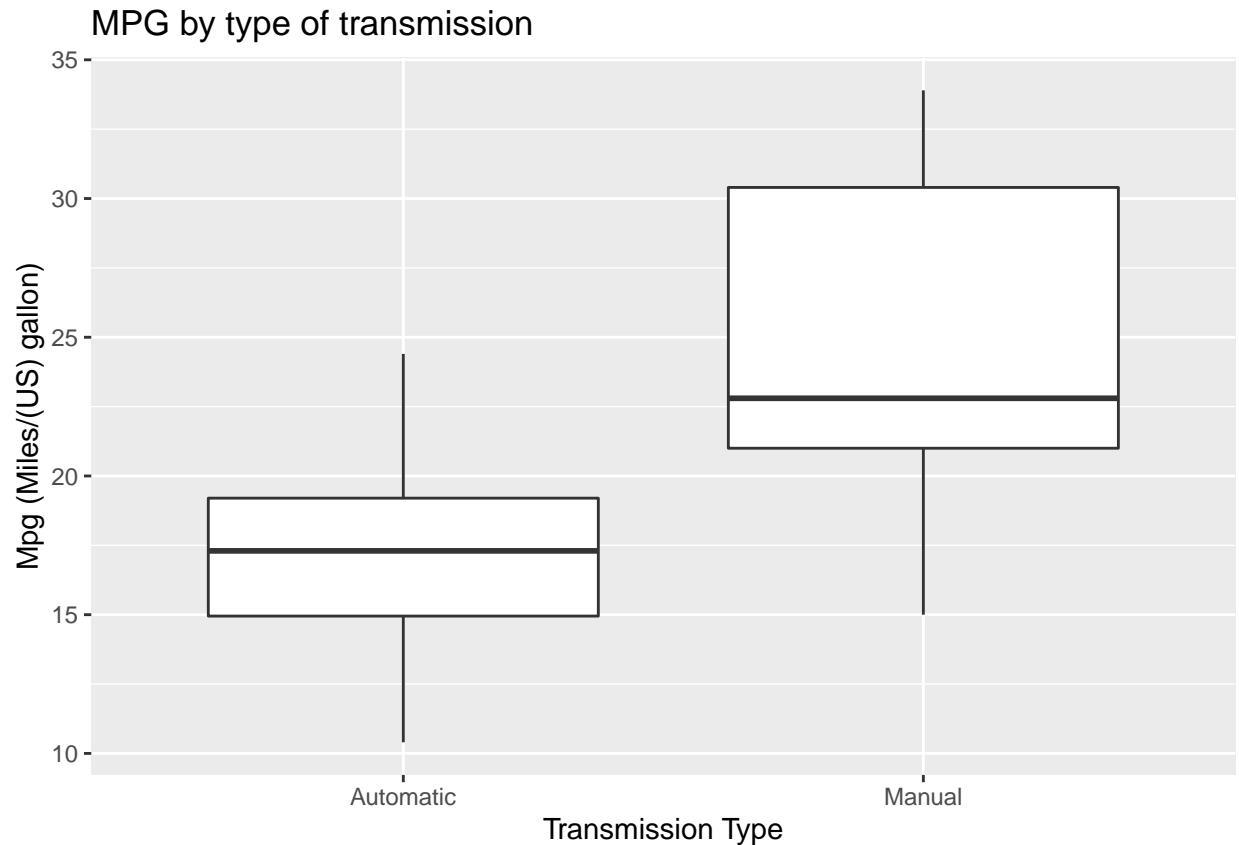
Manual

```
summary(data[data$am == 1, "mpg"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

Plot into ggplot

```
library(ggplot2)
ggplot(data, aes(x= factor(am, level=c(0,1), labels= c("Automatic","Manual")), y=mpg)) + geom_boxplot()
```

MPG by type of transmission

Multiple variable in models

Using step function to see variable impact by adding other variable

```r
par(mfrow =c(2,2))
fit <- step(lm(mpg ~ ., data))
```

```
## Start:  AIC=70.9
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##         Df Sum of Sq    RSS    AIC
## - cyl    1    0.0799 147.57 68.915
## - vs     1    0.1601 147.66 68.932
## - carb   1    0.4067 147.90 68.986
## - gear   1    1.3531 148.85 69.190
## - drat   1    1.6270 149.12 69.249
## - disp   1    3.9167 151.41 69.736
## - hp     1    6.8399 154.33 70.348
## - qsec   1    8.8641 156.36 70.765
## <none>               147.49 70.898
## - am     1   10.5467 158.04 71.108
## - wt     1   27.0144 174.51 74.280
##
## Step:  AIC=68.92
## mpg ~ disp + hp + drat + wt + qsec + vs + am + gear + carb
##
```

```
##         Df Sum of Sq    RSS    AIC
## - vs    1     0.2685 147.84 66.973
## - carb  1     0.5201 148.09 67.028
## - gear  1     1.8211 149.40 67.308
## - drat  1     1.9826 149.56 67.342
## - disp  1     3.9009 151.47 67.750
## - hp    1     7.3632 154.94 68.473
## <none>             147.57 68.915
## - qsec  1    10.0933 157.67 69.032
## - am    1    11.8359 159.41 69.384
## - wt    1    27.0280 174.60 72.297
##
## Step:  AIC=66.97
## mpg ~ disp + hp + drat + wt + qsec + am + gear + carb
##
##         Df Sum of Sq    RSS    AIC
## - carb  1     0.6855 148.53 65.121
## - gear  1     2.1437 149.99 65.434
## - drat  1     2.2139 150.06 65.449
## - disp  1     3.6467 151.49 65.753
## - hp    1     7.1060 154.95 66.475
## <none>             147.84 66.973
## - am    1    11.5694 159.41 67.384
## - qsec  1    15.6830 163.53 68.200
## - wt    1    27.3799 175.22 70.410
##
## Step:  AIC=65.12
## mpg ~ disp + hp + drat + wt + qsec + am + gear
##
##         Df Sum of Sq    RSS    AIC
## - gear  1      1.565 150.09 63.457
## - drat  1      1.932 150.46 63.535
## <none>             148.53 65.121
## - disp  1     10.110 158.64 65.229
## - am    1     12.323 160.85 65.672
## - hp    1     14.826 163.35 66.166
## - qsec  1     26.408 174.94 68.358
## - wt    1     69.127 217.66 75.350
##
## Step:  AIC=63.46
## mpg ~ disp + hp + drat + wt + qsec + am
##
##         Df Sum of Sq    RSS    AIC
## - drat  1      3.345 153.44 62.162
## - disp  1      8.545 158.64 63.229
## <none>             150.09 63.457
## - hp    1     13.285 163.38 64.171
## - am    1     20.036 170.13 65.466
## - qsec  1     25.574 175.67 66.491
## - wt    1     67.572 217.66 73.351
##
## Step:  AIC=62.16
## mpg ~ disp + hp + wt + qsec + am
##
```

```
##         Df Sum of Sq    RSS    AIC
## - disp  1      6.629 160.07 61.515
## <none>              153.44 62.162
## - hp    1     12.572 166.01 62.682
## - qsec  1     26.470 179.91 65.255
## - am    1     32.198 185.63 66.258
## - wt    1     69.043 222.48 72.051
##
## Step:  AIC=61.52
## mpg ~ hp + wt + qsec + am
##
##         Df Sum of Sq    RSS    AIC
## - hp    1      9.219 169.29 61.307
## <none>              160.07 61.515
## - qsec  1     20.225 180.29 63.323
## - am    1     25.993 186.06 64.331
## - wt    1     78.494 238.56 72.284
##
## Step:  AIC=61.31
## mpg ~ wt + qsec + am
##
##         Df Sum of Sq    RSS    AIC
## <none>              169.29 61.307
## - am    1     26.178 195.46 63.908
## - qsec  1    109.034 278.32 75.217
## - wt    1    183.347 352.63 82.790
```

```r
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

```r
plot(fit)
```
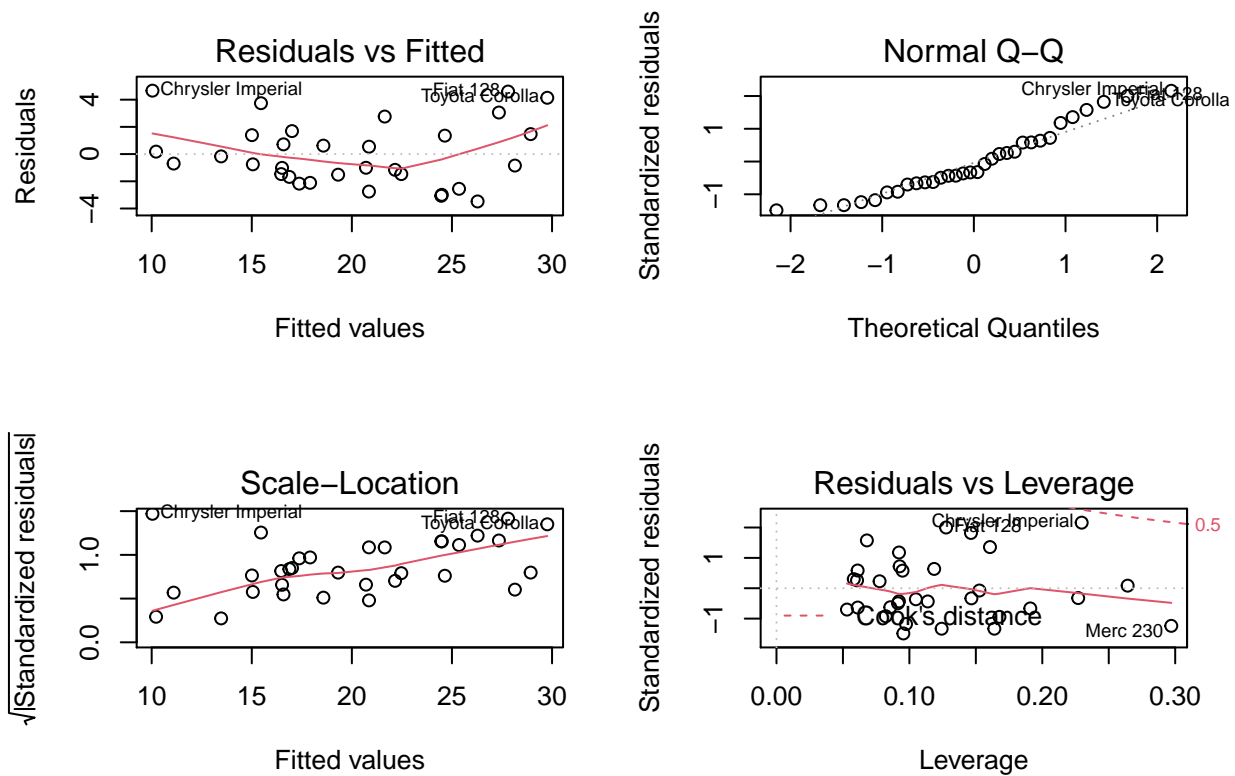
Conclusions

Through analyzing the mtcars dataset, we found that manual transmission cars have more mpg (miles per gallon) than automatic transmission cars.

We also found that wt/qsec/am predictor is the best model. Multiple R-squared values for this model are 84.97% which is high enough for mpg outcome.