Relationship between mpg and transmission

DBisure

28/07/2020

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:

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

Processing the data

```
summary(mtcars)
```

Exploration

Checking the corelation of mpg with other variables

```
motorData <- mtcars
cor(motorData$mpg,motorData[,-1])
```

```
##            cyl       disp         hp      drat         wt     qsec        vs
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684 0.6640389
##             am      gear       carb
## [1,] 0.5998324 0.4802848 -0.5509251
```

We notice that cyl,disp,hp,wt, carb have a negative co relation with mpg. That means if we keep all other variables constant, With increase in cyl(example) we should see a decrease in mpg. A plot is present in Appendix 2

Automatic v/s Manual Transmission

In the dataset, "am" tells us about the transmission, where 0 = automatic and 1 = manual. Changing numeric vector into a factor vector. The relationship between transmission and mpg can be in seen in plot in Appendix 1

```r
motorData$am <- as.factor(motorData$am)
levels(motorData$am) <-c("Automatic", "Manual")
```

Performing a t-test

We will perform a t-test to find any signficant difference between manual and auto transmission

```r
t.test(motorData$mpg~motorData$am,conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  motorData$mpg by motorData$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 0.001374 hence we can reject the hypothesis that mean mpg of automatic and manual is the same, we find out that automatic transmission has lower mpg as compared to manual. However, this is with the fact that all other variables are kept constant.

Finding effect of other variables on mpg

```r
multivariate_model <- lm(data=motorData, mpg ~ .)
summary(multivariate_model)$coefficients[,1:4]
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## amManual     2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

wt has a lot of effect on the mpg, we will try to find a model which best suits the data and explains most of the variance, we will use the step function

```r
global_multivariate <- step(lm(data=motorData, mpg ~ .), trace=0)
summary(global_multivariate)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = motorData)
##
## 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 ***
## amManual      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
```

The results suggests that the best model includes qsec, wt, and amManual variables. About 85% of the variance is explained by this model. Weight change negatively with mpg around 3.9165miles/ galon every 1000lbs. Manual transmission is 2.9mpg better than automatic transmission.

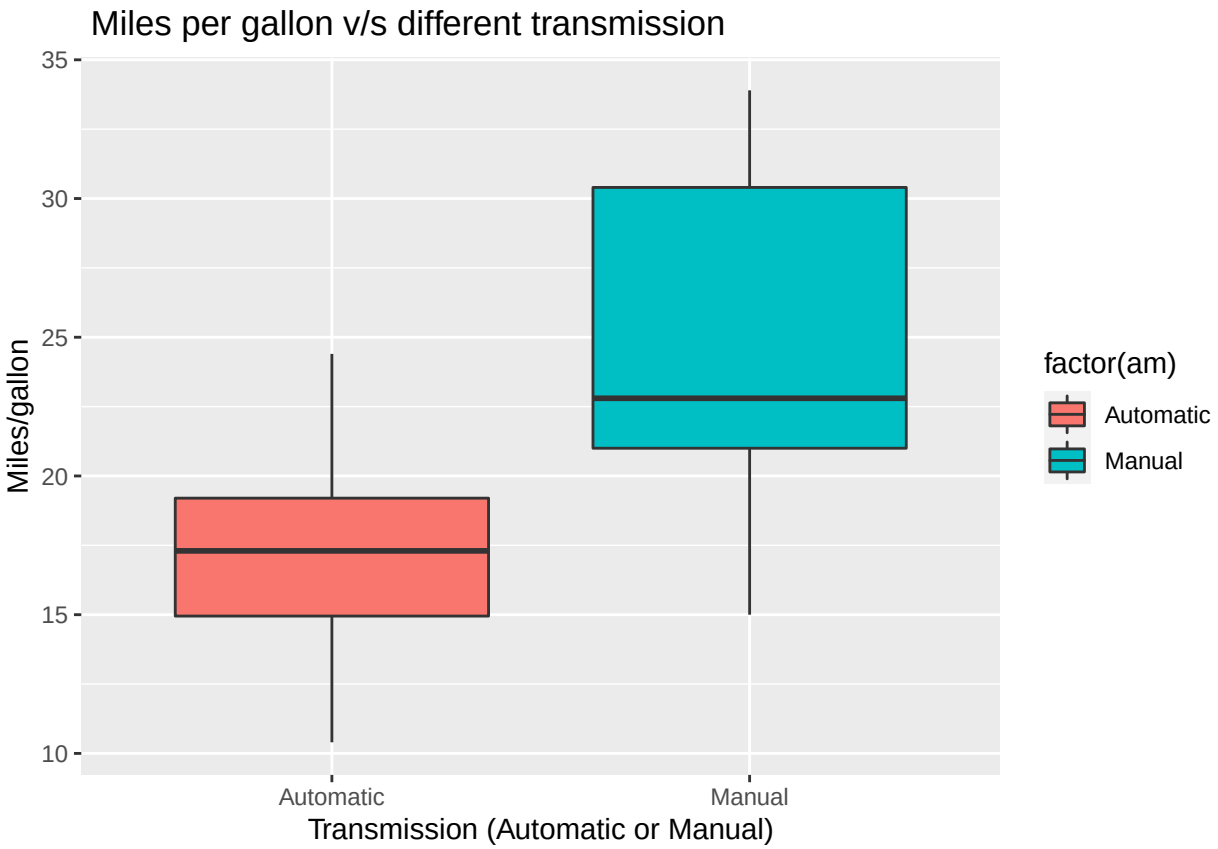
Conclusion

On average, manual transmission is better than automatic transmission by 2.9 mpg. However, transmission type is not the only factor accounting for mpg - weight,qsec are also affecting the MPG.

Appendix: Figures

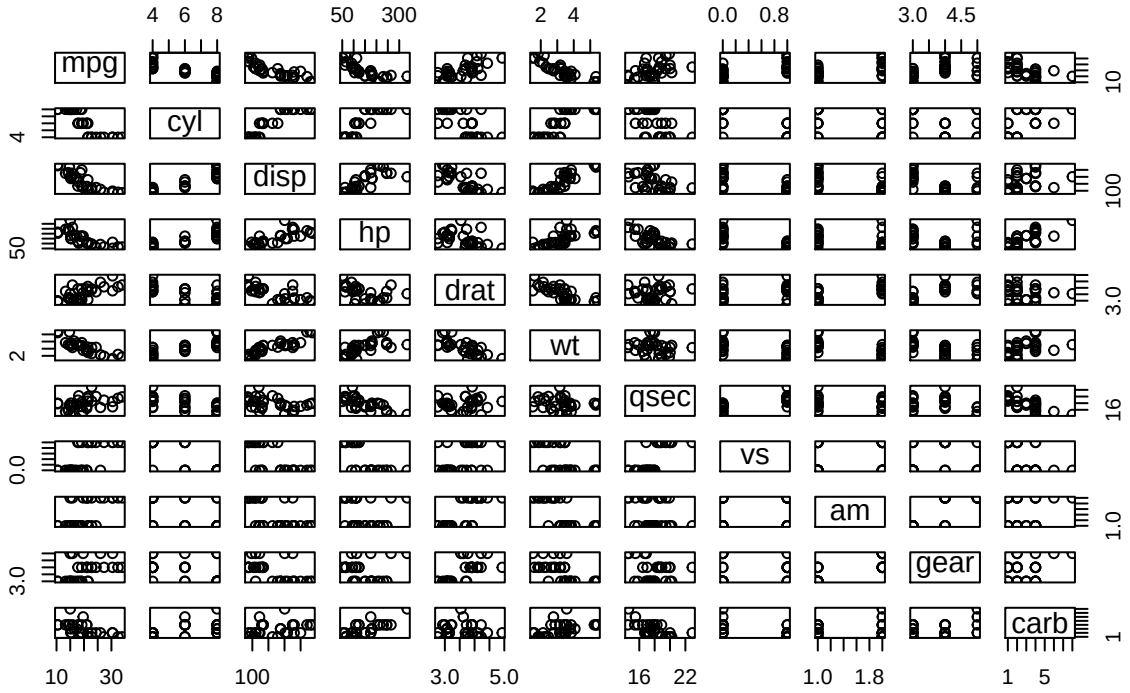
Appendix 1 : Plot for different Transmissions

```
library(ggplot2)
g <- ggplot(motorData, aes(x=factor(am), y=mpg, fill=factor(am)))
g + geom_boxplot(notch=F) +
       scale_x_discrete("Transmission (Automatic or Manual)") +
       scale_y_continuous("Miles/gallon") +
       ggtitle(" Miles per gallon v/s different transmission")
```



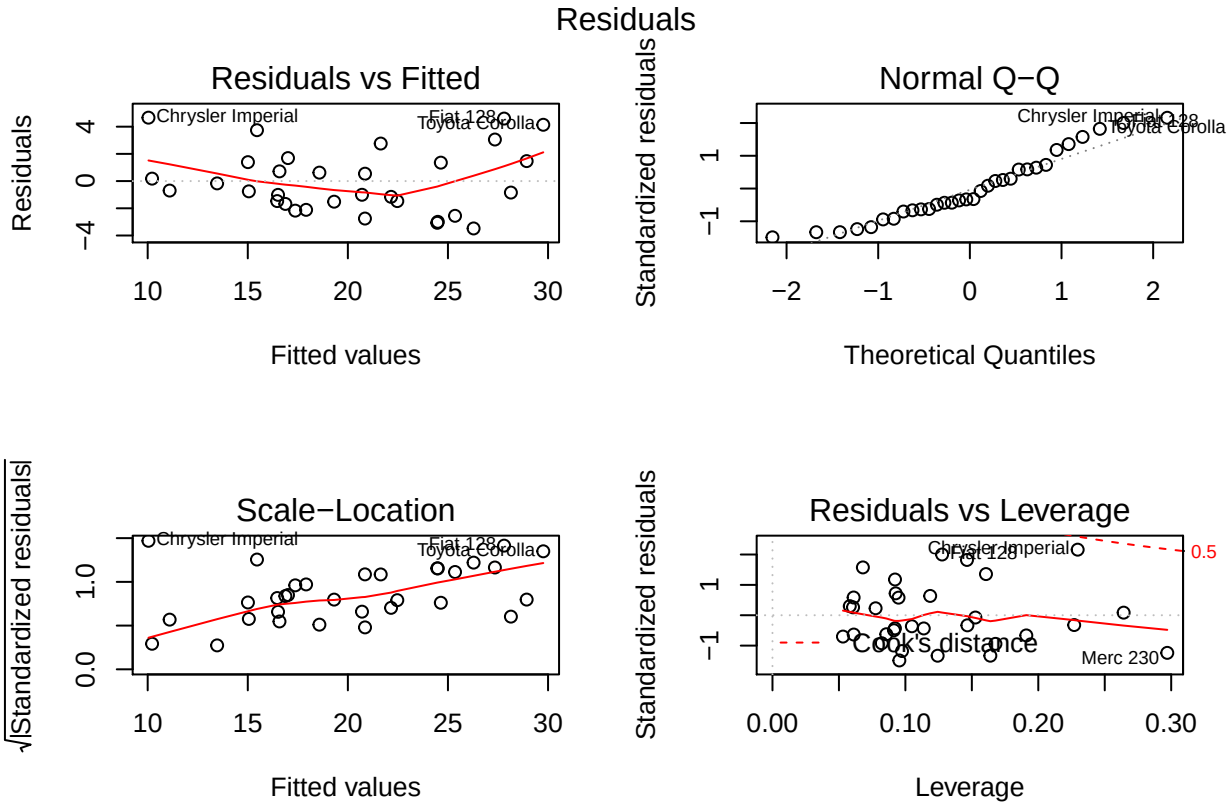
Appendix 2: Matrix

```
pairs(mpg ~ ., data = motorData, main="Relationship between all the variables")
```



Appendix 3: Residual Plot

```
par(mfrow = c(2,2))
plot(global_multivariate)
mtext("Residuals", side = 3, line = -2, outer = TRUE)
```


Residuals
Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Cook's distance
0.5
Merc 230
Leverage