Automatic vs Manual Transmission in term of MPG

Ameen AboDabash

Sep 2, 2016

Executive Summary

We'll explorer *mtcars* dataset to build a regression model which will confirm Manual is better than Automatic transitions in term of MPG "Miles Per Gallon" (avg 2.9 mpg much better), and quantify that difference.

Setting the Scene

Let's look at the data first, *mtcars* 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), So *mtcars* is A data frame with 32 observations on 11 variables (Fig.1 Scatter plot for all vars)

[, 1] mpg *Miles/(US) gallon*	[, 9] am *Transmission (0=automatic, 1=manual)*
[, 2] cyl *Number of cylinders*	[, 3] disp *Displacement (cu.in.)*
[, 4] hp *Gross horsepower*	[, 5] drat *Rear axle ratio*
[, 6] wt *Weight (lb/1000)*	[, 7] qsec *1/4 mile time*
[, 8] vs *V/S*	[,10] gear *Number of forward gears*
[,11] carb *Number of carburetors*	.

Frankly speaking, its seems Manual transmission better than Automatic from data plotting "Appendix Fig.1", So lets get the *P-Value* to confirm this "we can call it" assumption or hypothesis:

```
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
```

So, P-Value is 0.001374 which s less than 0.05 which confirm our first assumption.

Regression Model and Qunitifiying the difference

So the following basic model which multivariate linear regression with all variables:

```
summary(lm(mpg~am,mtcars))
```

```
##
```

```
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## amManual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Indeed, it looks like wt is the only factor that significantly changes with mpg. However, including all variables will possibly result over-fitting, R has an automatic variable-selection function *step*. so lets enhance our model using *step* as following:

```
summary(step(lm(data = mtcars, mpg ~ .), trace=0))
```

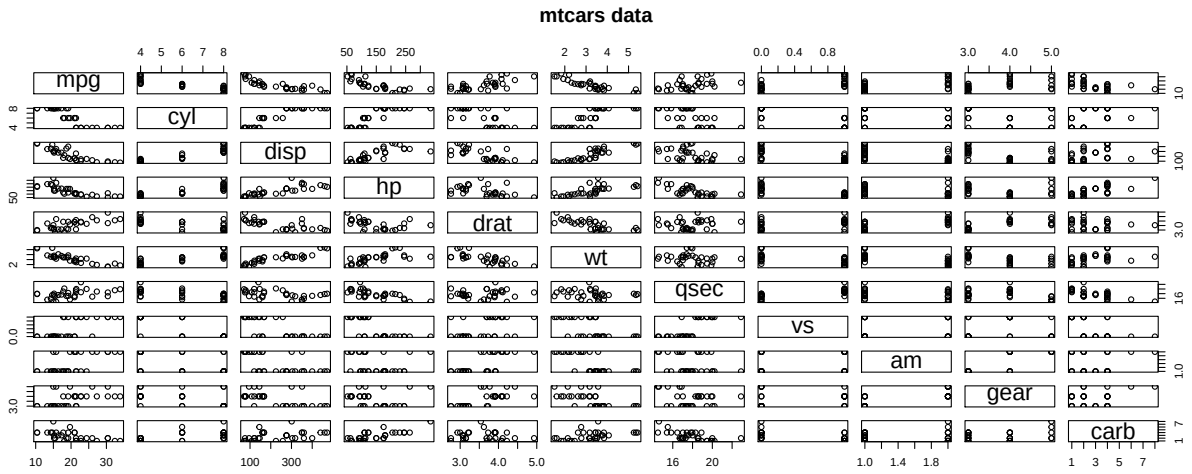
```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -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
```

So *wt (weight)* included by step function and *qsec (1/4 mile time)* as long as *am (Transimession Type)* , *wt* negatively changes with *mpg*, and *qsec* and *am* positively changes. Every lb/1000 weight increase will cause a decrease of roughly 4 mpg, every increase of 1/4 mile time will cause an increase of 1.2 mpg, and on average, manual transmission is *2.9 mpg better* than automatic transmission.Appendix(Fig.2).

Appendix

Fig.1

```
require(graphics)
pairs(mtcars, main = "mtcars data")
```




```
#coplot(mpg ~ am | as.factor(am), data = mtcars, panel = panel.smooth, rows = 1)
boxplot(mpg~am, data = mtcars,
        xlab = "Transmission",
        ylab = "MPG",
        main = "MPG in term of MPG")
```

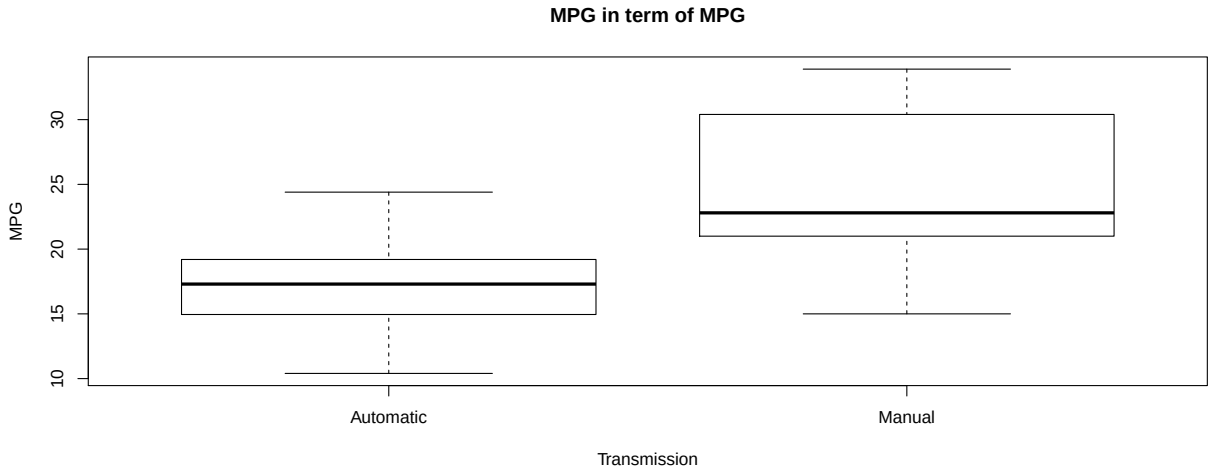
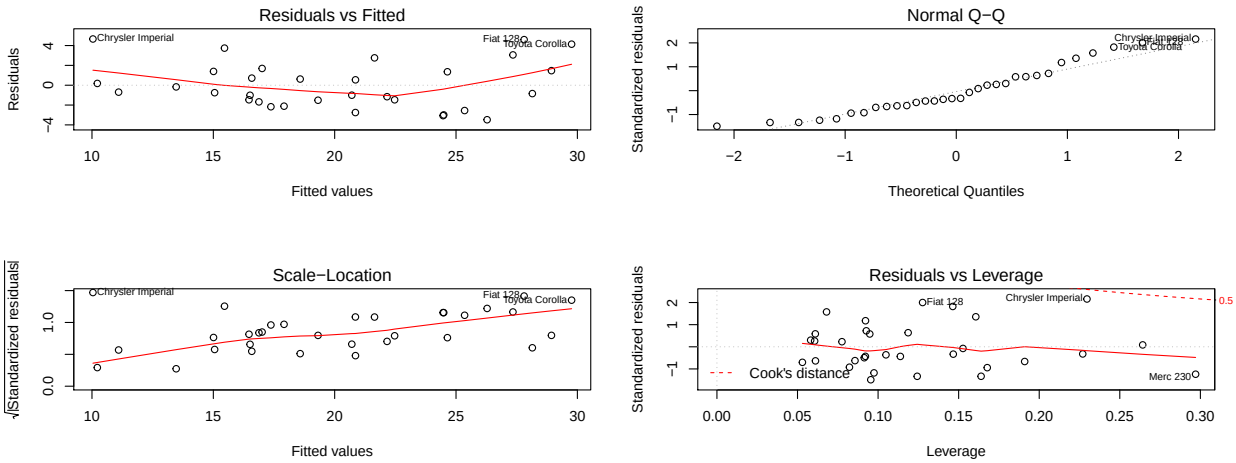


Fig.2

```
par(mfrow = c(2,2))
plot(step(lm(data = mtcars, mpg ~ .), trace=0))
```



Analysis available on [github]https://github.com/aabodabash/autoVsManualMPG.git