Motor-Trends

Fernando Campaña

10/13/2019

Relationship between MPG and Transmission

In the following report we are going to explore the relationship between Miles per Gallon (MPG) and cars transmissions (automatic or manual), by fitting different models in order to answer the following questions:

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

To answer the questions we are going to prove if one transmission is better than the other with a t-test, to further proceed with a model fitting of a simple linear regression and a multivariate linear regression.

Explorating the data

The data in analysis is the mtcars dataset available in R by default. The data contains a data frame with 32 observations on 11 (numeric) variables:

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

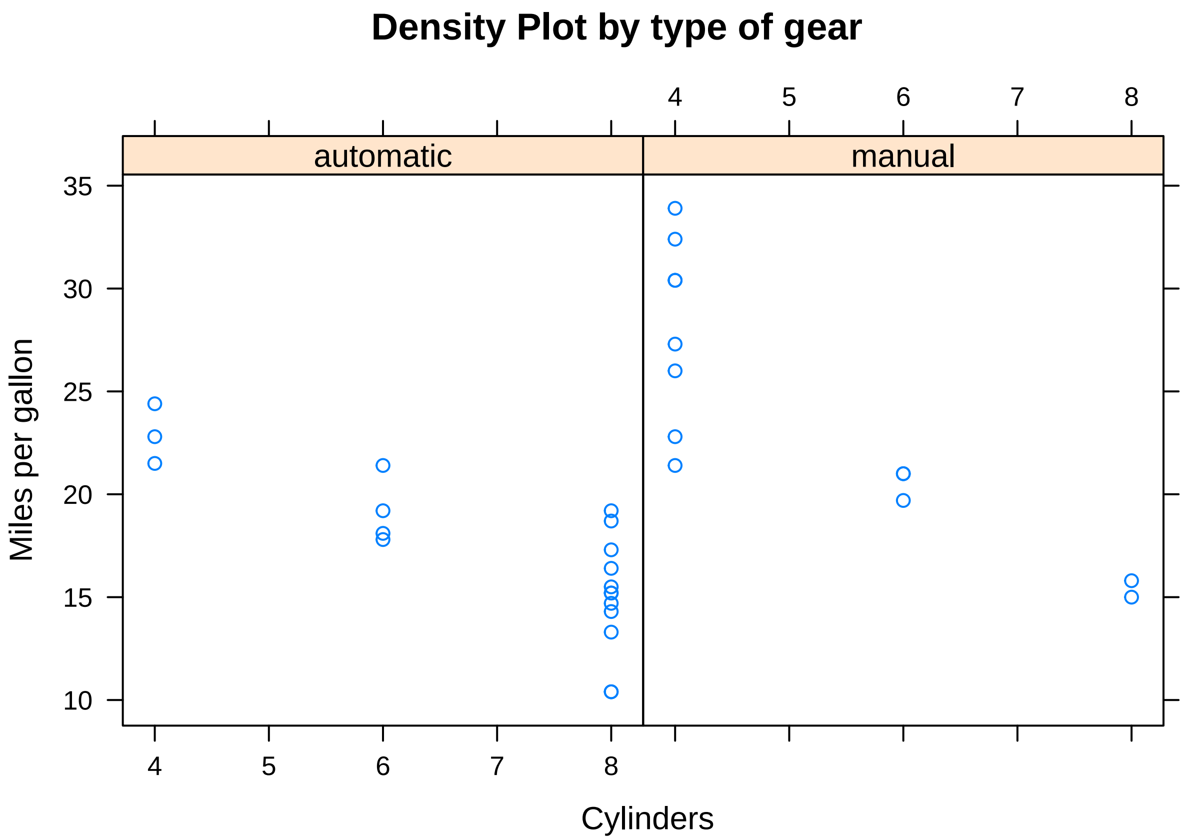
We print the first five values to visualize a subset of the data.

```r
kable(head(mtcars, 5))
```

	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

Then, we visualize the data with a plot:

```r
am.f<-factor(am, levels=c(0,1),
   labels = c("automatic","manual"))
xyplot(mpg ~ cyl | am.f,
   main = "Density Plot by type of gear",
   xlab =" Cylinders", ylab = "Miles per gallon")
```



It looks that automatic transmission can travel less miles per gallon than the manual, especially for 8 cylinders. It seems to be a trend that more cylinders gives less miles per gallon, and the conversely could also be true. We can test it through a T-test:

```
# Transform the numerical factor into character labels
mtcars$am <- factor(mtcars$am, labels=c('Automatic','Manual'))
t.test(mpg ~ am,  mtcars, paired = FALSE, var.equal = FALSE,conf.level = .95)
```

```
##
##  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
```

Because the p-value is so small we can be confident that the Manual transmission is better for mpg than the Automatic, with a higher mean of 24.39. So the manual cars can cover more miles with the same gallons than automatic cars.

Now, we will quantify the answer by fitting two models, a simple linear regression and a multivariate regression.

Simple Linear Regression

To quantify the mpg difference we would like a model like mpg = f(am), which we can obtain by fitting a linear regresion:

```
simplefit <- lm(mpg ~  am, data = mtcars)
summary(simplefit)
```

```
##
## 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
```

The result gives us the model: mpg = 17.147 + 7.245 amManual

Multivariate Linear Regression

Given that the dataset has many variables, somo can be very correlated which will add variance, others could have little effect on the model. So we can use a AIC step-wise algorithm that will produce various models with some variables taken in account and automatically select the best one which has the least variance.

```
multifits <- lm(mpg ~  ., data = mtcars)
multifit <- (step(multifits, trace=0, direction = "backward"))
```

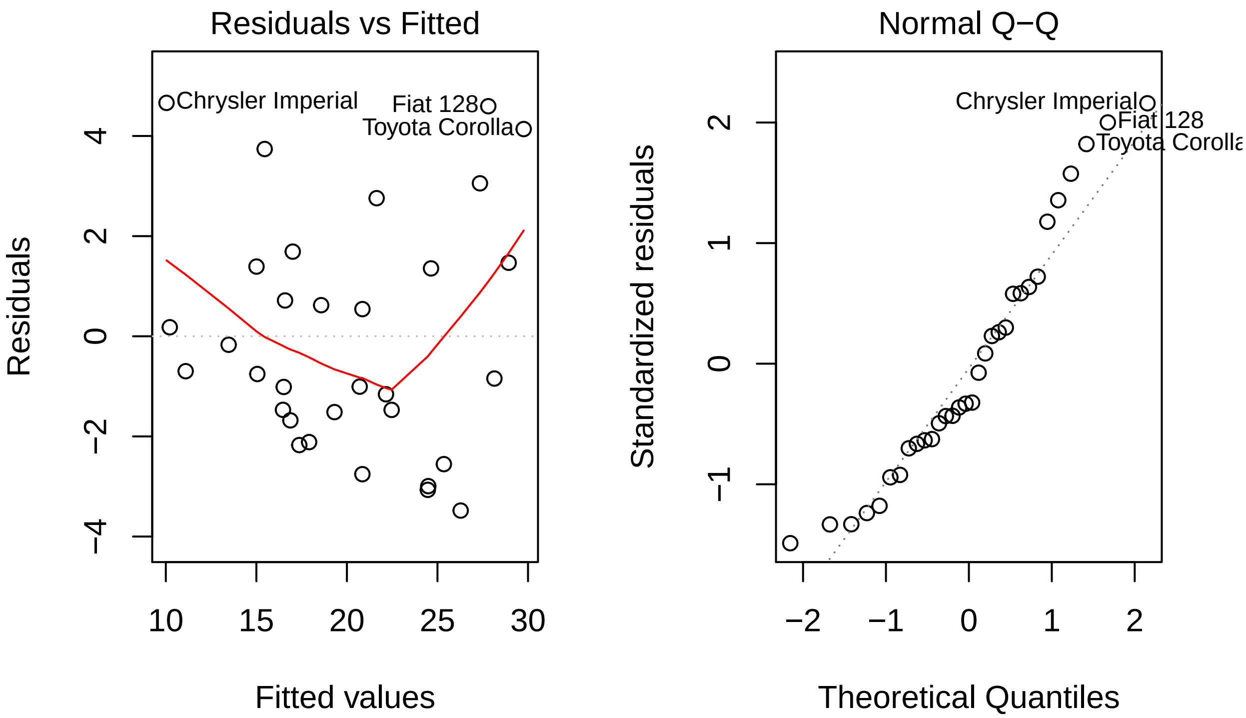
```
summary(multifit)
```

```
##
## 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
```

The fitted model shows us that the most useful variables for our question are wt (weight), qsec (1/4 mile time) and am (transmission), giving a formula like mpg = 9.6178 - 3.9165 wt + 1.2259 qsec + 2.9358 am.

```r
par(mfrow = c(1,2))
plot(multifit, which = c(2,1))
```



As we can see from the diagnostic plots, the normalized residuals lie on the normal line (Normal Q-Q) and the residuals are randomly distributed so it doesn't look biased and confirms that the model is efficient.

Conclusion

- Is an automatic or manual transmission better for MPG?

A manual transmission is better for mpg than an automatic transmission, proved by a T-test with a p-value of 0.001374.

- Quantify the MPG difference between automatic and manual transmissions?

By looking at the multivariate fitted model: mpg = 9.6178 - 3.9165 wt + 1.2259 qsec + 2.9358 amManual. This tells us that the amManual has a 2.9358 effect compared to other variables, less than the 7.245 of the simple linear model.

Compared to the simple linear regression with a R-squared of 0.3598 while the multivariate regression gives an R-squared of 0.8497, the higher R-squared tells us that the multivariate has less difference between the fitted model and the data with a 85% of the variance of mpg.