Comparison of Manual and Automatic Transmission in Consideration of MPG

Oliver Stroemer

2019-07-22

Executive Summary

In this report we're going to investigate the question if an manual or an automatic transmission is better for the miles per gallon quota and if, how big the effect is.

We conclude that, yes, using a manual transmission lets us expect about 5.28 mpg more ($p < 0.001$).

Load the data

We'll use the `mtcars` data set from the R `datasets` package.

```
data(mtcars)
```

Exploration

Detailed information about this data set can be obtained with the command `?mtcars`. The MPG is stored in the `mpg` column. The `am` column contains the type of the transmission: `0` for automatic and `1` for manual.

We should get a quick overview over the relation of `mpg` to `am`. Let's have a glance at box plot.

```
boxplot(mpg ~ am, data=mtcars)
```

You can find the box plot in the appendix section of this report. Indeed it seems as if there were a correlation between the type of transmission and `mpg`.

Fitting models

Now we try to proof our claim using different linear regression models. We assume that `mpg` follows approximately a normal distribution; `mpg` can't be negative, therefore it can't be normal.

Model 1: Using only type of transmission as predictor

In this first model, we'll only use the type of transmission as a predictor. Although it isn't necessary to convert `am` to factor, since it already is `0` or `1`, we'll do it because of the semantics.

```
fit1 <- lm(mpg ~ factor(am), mtcars)
summary(fit1)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fit1)$adj.r.squared
```

```
## [1] 0.3384589
```

As we've already guessed there is a very significant correlation. `mpg` increases by about 7.24 mpg if the car has an manual transmission ($p < 0.001$).

But this model accounts for only about 34 % of the variance. We should do some residual diagnostics (the plot is again in the appendix):

```
par(mfrow=c(2,2))
plot(fit1)
```

That doesn't seem right. The residuals are split in two groups. Since we accounted for only about a third of the variance it isn't unlikely, we missed a confounding variable in our model.

Model 2: Including horse power in the model

Let's try to include the `hp` column of our data in our model. Horse power might be strongly correlated with mpg, too, and therefore might confound our relationship between mpg and transmission type.

```
fit2 <- lm(mpg ~ factor(am) + hp, mtcars)
summary(fit2)$coef
```

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
## factor(am)1  5.2770853 1.079540576  4.888270 3.460318e-05
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
```

```
summary(fit2)$adj.r.squared
```

```
## [1] 0.7670025
```

Gotcha! There is a highly significant relationship between `mpg` and `hp`. Now we account for about 77 % of the variance. Let's have again a look at our residuals:

```
par(mfrow=c(2,2))
plot(fit2)
```

That looks way better. Although there seems to be some room for improvement, the residuals spread without big anomalies.

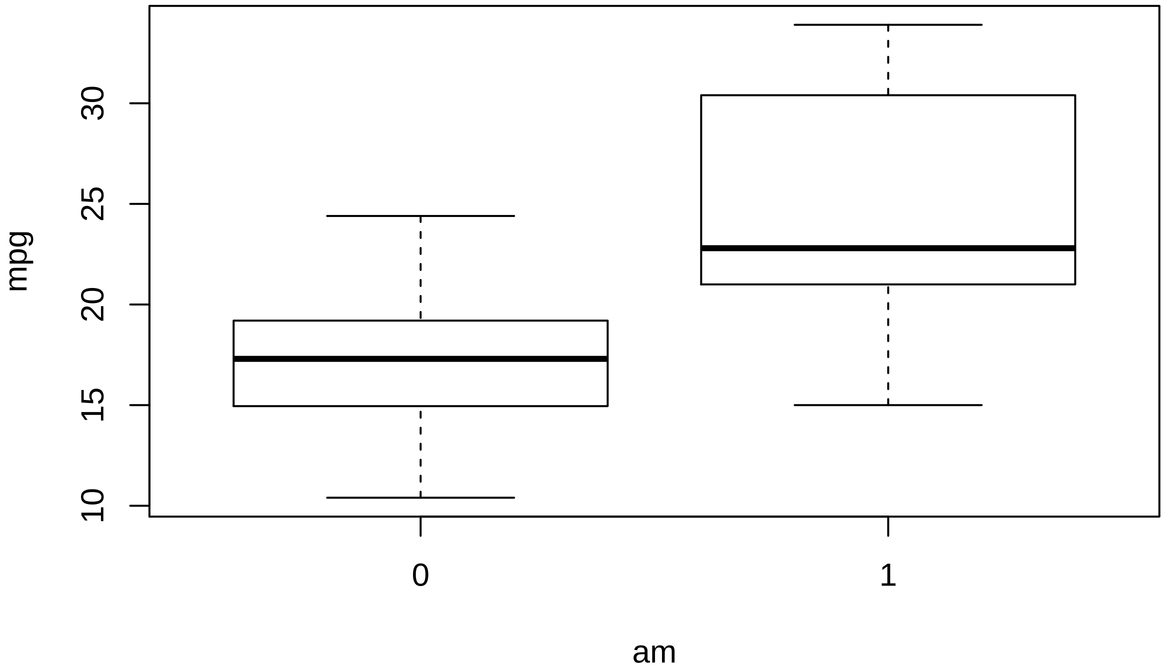
Conclusions

Whether or not we use a manual transmission impacts highly our mpg. Using a manual transmission we can expect an additional 5.28 mpg bonus ($p < 0.001$) even if we include horse power as a confounder in our model. With each increment of `hp` we can expect to have 0.06 mpg less ($p < 0.001$).

Appendix

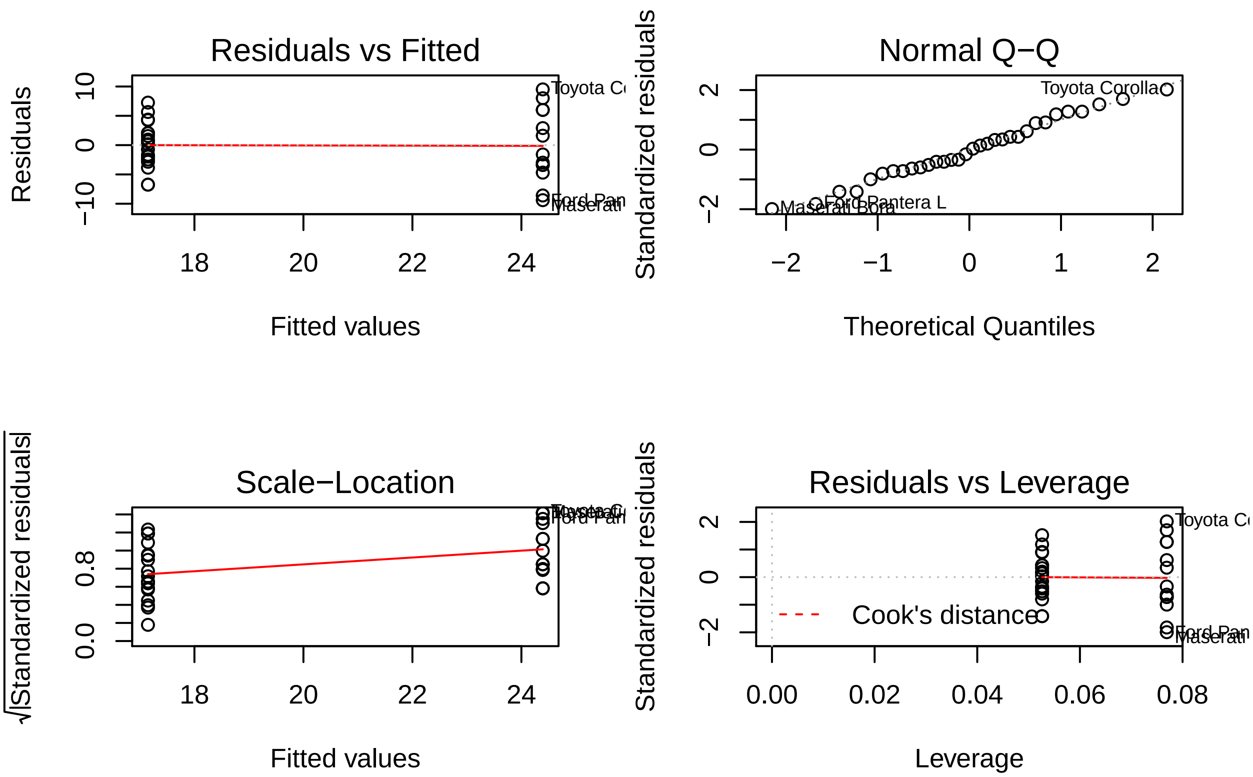
Exploration of the relationship between `mpg` and `am`

```
boxplot(mpg ~ am, data=mtcars)
```



Residuals Model 1

```
par(mfrow=c(2,2))
plot(fit1)
```



Residuals Model 2

```
par(mfrow=c(2,2))
plot(fit2)
```