Motor Trend analysis: Automatic or manual transmission?

Daria Stepanyan

Executive summary

This report contains analysis of set of cars data for Motor Trend, a magazine about the automobile industry. It's dedicated to exploring the relationship between transmission types and miles per gallon (MPG) (outcome), particularly the following two questions:
1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.
Regression analysis was performed, and the results obtained show that factors such as number of cylinders, horsepower and weight mostly affect mpg, and the influence of transmission type is not statisctically significant.

Exploratory analysis

```
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

The 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). Dataset description can be found here: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html

Let's take a look at the rest of the variables relations to mpg.

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

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

From the correlation results, we can see, that the cyl, hp, wt and carb are negatively correlated with mpg.

Transmission data is stored in variable "am" (0 = automatic, 1 = manual). For easier interpreptation, let's convert it to the factor variable.

```
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

The boxplot (appendix 1) shows that cars with manual transmission have better mpg (can go more distance on the same amount of gas). When plotting the data across different cylinders amount though, the result is

not as straigntforward (Appendix 2).

Regression analysis

Let's see how much of the variance in mpg is actually explained by transmission type.

```
summary(lm(mpg~ am,data=mtcars))$r.squared
```

```
## [1] 0.3597989
```

While manual transmission cars seem to have more mpg on average, with this variability we can't say that transmission type is the only influencing parameter. This needs to be further explored in a multiple linear regression analysis.

We use the step function in R for a stepwise regression, where the choice of predictor is carried out automatically.

```
best = step(lm(data = mtcars, mpg ~ .),trace=0)
best$call
```

```
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
```

```
summary(best)$r.squared
```

```
## [1] 0.8496636
```

Interestingly, the qsec is chosen as a defining parameter, since it clearly depends on car's design and such parameters as horse power and number of cylinders. Let's see if the model will be better fitted if we use those instead of their dependant, qsec.

```
final <- lm(mpg~ am+wt+hp+cyl,data=mtcars)
summary(final)$r.squared
```

```
## [1] 0.8490314
```

This model explains almost the same amount of variance in mpg, but is way easier for interpretation, so let's assume it as our final model.

```
summary(final)$coefficients
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 36.14653575 3.10478079 11.642218 4.944804e-12
## amManual     1.47804771 1.44114927  1.025603 3.141799e-01
## wt          -2.60648071 0.91983749 -2.833632 8.603218e-03
## hp          -0.02495106 0.01364614 -1.828433 7.855337e-02
## cyl         -0.74515702 0.58278741 -1.278609 2.119166e-01
```

We can see that with all other parameters being equal manual transmission increases mpg by 1.4780477. Residuals (see Appendix 3) are also randomly scattered.

Inference

However with 95% confidence we can't say that manual transmission always has a positive effect on mpg, so the influence of transmission type is not statistically sinificant.

```
confint(final, 'amManual')
```

```
##              2.5 %   97.5 %
## amManual -1.478946 4.435042
```

We can also check if our model is significant, in other words is a better fitted than a model with just am as predictor.

```r
anova(lm(mpg~ am,data=mtcars), final)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + hp + cyl
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.9
## 2     27 170.0  3     550.9 29.166 1.274e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is significant, we reject the null hypothesis that the models are similar and conclude that the additional variables in our final model are necessary.

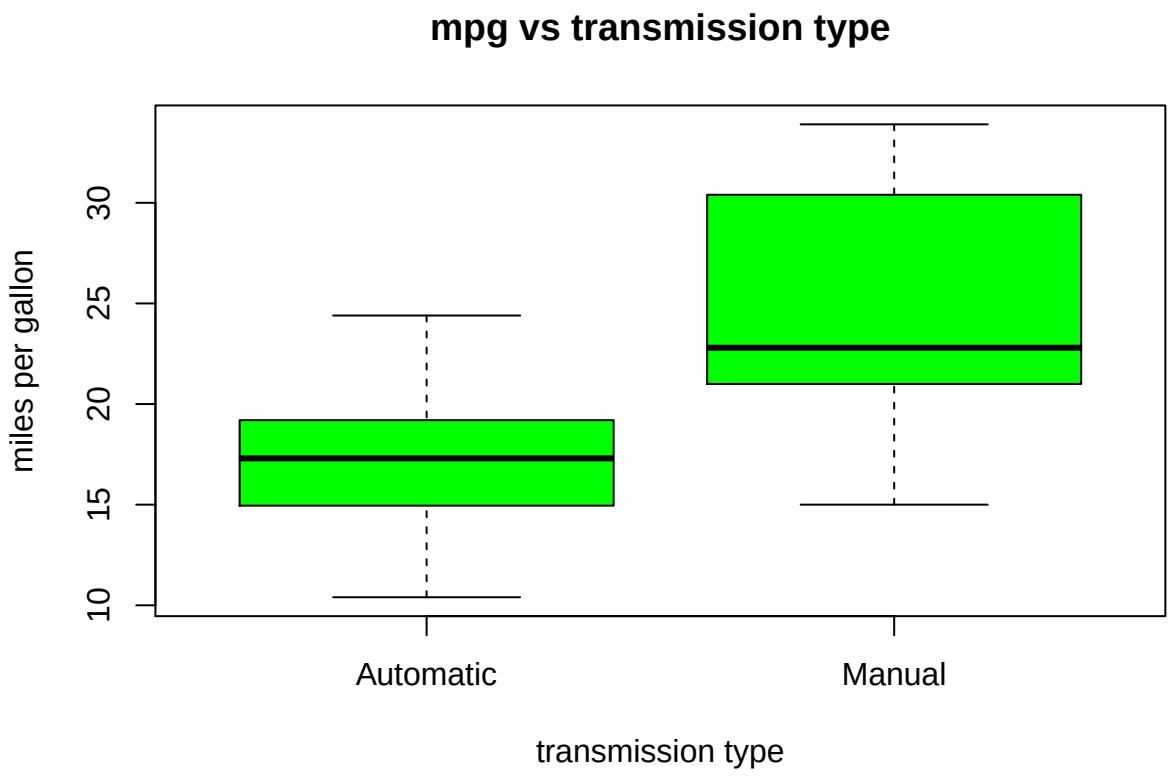
Final conclusion

According to the above research, transmission type is not the most influential parameter for the final mpg value (weight, horsepower and number of cylinders should be considered primarily). And although, according to the model, choosing manual transmission over automatic with all the rest being equal can increase mpg by 1.4780477, those results are not significant enough and more studies for each car design configuration must be conducted.

Appendix

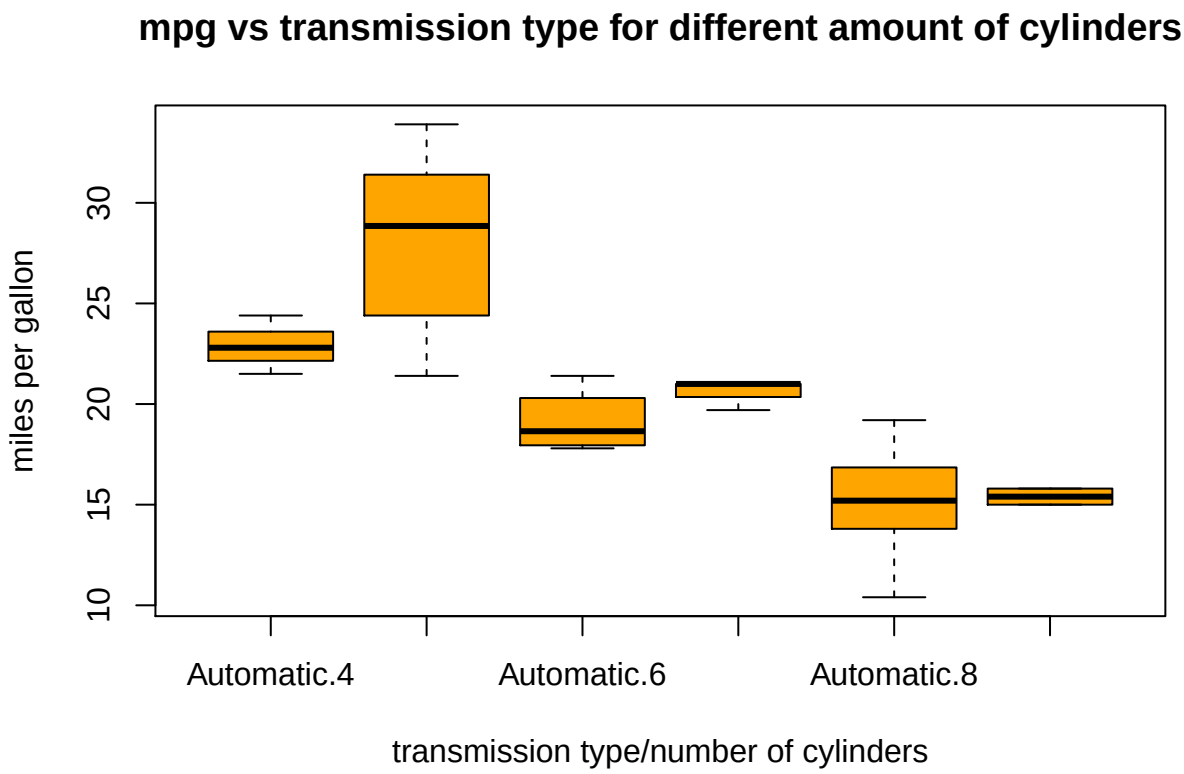
Apendix 1. Boxplot of mpg per transmission type

```r
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars,
        ylab="miles per gallon",
        xlab="transmission type",
        main="mpg vs transmission type",
        col="green")
```



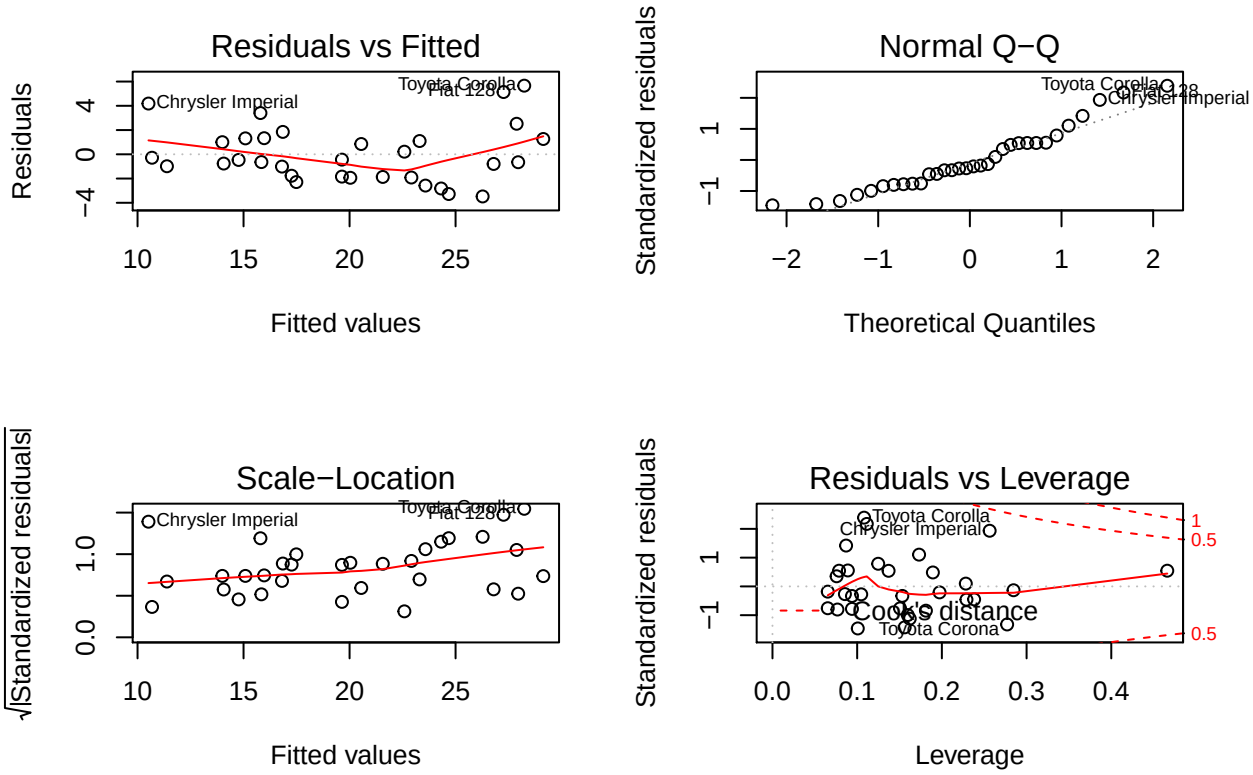
Appendix 2. Boxplot of mpg per transmission type per cylinder

```
boxplot(mtcars$mpg ~mtcars$am + mtcars$cyl, data = mtcars,
        ylab="miles per gallon",
        xlab="transmission type/number of cylinders",
        main="mpg vs transmission type for different amount of cylinders",
        col="orange")
```



Appendix 3. Residuals

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


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