# Effect of transmission type to fuel economy

*Martí Cuquet*

## Executive summary

In this report we use the Motor Trend Car Road Tests (`mtcars`) dataset to explore the relation between 10 different aspects of automobile design and performance with the fuel economy, with a special focus on the effect of the transmission type (either automatic or manual). The dataset, available from the `datasets` R package, was extracted from the 1974 *Motor Trend* US magazine and contains 32 observations. We conclude that the transmission type has indeed an effect to the fuel economy, but that this effect is lower if one accounts for other variables such as weight or acceleration of the car.

## Exploratory data analysis

First, we load the `mtcars` dataset and convert some of the variables to factors.

```
library(dplyr); library(ggplot2); library(GGally); library(pander); library(tidyr)
data(mtcars); corr.matrix <- mtcars
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs, levels=c(0, 1), labels=c("V-engine", "straight engine"))
mtcars$am <- factor(mtcars$am, levels=c(0, 1), labels=c("automatic", "manual"))
mtcars$gear <- factor(mtcars$gear); mtcars$carb <- factor(mtcars$carb)
```

The first impression is that the type of transmission has a big impact in the expected fuel economy.

```
g.box.am <- ggplot(mtcars, aes(am, mpg, fill=am)) + geom_boxplot() + xlab("") +
  ylab("Fuel economy (mpg)") + guides(fill=guide_legend(title="Type of transmission"))
```

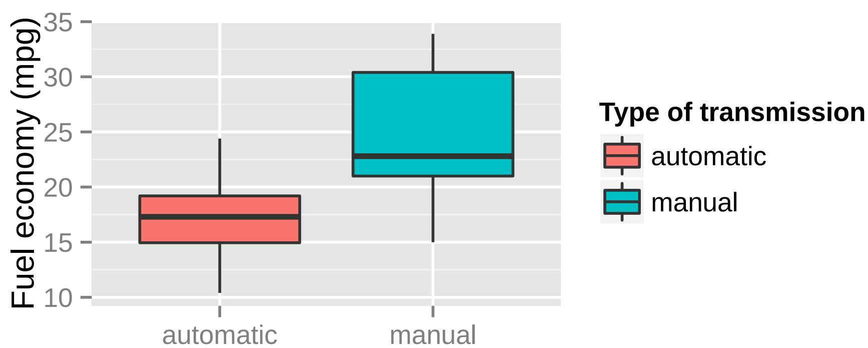


Figure 1: Fuel economy depending on the transmission type in the `mtcars` dataset.

We can indeed test that the two transmission types have a different effect on the fuel economy. Assuming roughly symmetric and mound shaped distribution, we perform a $t$-test with the alternative hypothesis that the true difference in means is different from 0.

```
t <- t.test(mpg ~ am, data=mtcars)
```

The 95% confidence interval lies between (-11.28, -3.21) and the $p$-value is 0.0013736, so we can reject the null hypothesis that the transmission type has no effect on the fuel economy.

The correlation matrix can be seen in Fig. 2 (in Appendix A). We can observe that the transmission type is indeed quite correlated with the fuel economy (0.60), but there are other variables that are even more correlated with it, and that are still somehow correlated with the transmission type (see for example the weight, with a correlation of -0.87 with the fuel consumption and of -0.69 with the transmission type). The panel in Fig. 3 (in Appendix A) shows the relation between fuel economy and the four variables most correlated with it.

## Models

In this section we explore several linear regression models. We first explore a model that considers only the transmission type as the single independent variable.

```
model.am <- lm(mpg ~ am, data=mtcars)
```

The model predicts that a manual transmission increases the fuel economy by 7.245 with respect to an automatic transmission, as could be already observed in Fig. 1. It has an adjusted $R^2$ value of 0.338, and both the intercept and the transmission type are significant above 0.05 (see Table 1 in Appendix B for more details). Considering a model with all possible variables,

```
model.full <- lm(mpg ~ ., data=mtcars)
```

a manual transmission increases the fuel economy by only 1.212 with respect to an automatic transmission. In this case, however, even if the adjusted $R^2$ is 0.779, none of the variables has a significance above 0.05. We hence use the `step` function to select the optimal model by the AIC.

```
model.aic <- step(model.full)
```

As we can see in Table 2 (in Appendix B), now a manual transmission increases the fuel economy by 1.809 with respect to an automatic transmission. The adjusted $R^2$ value is in this case 0.84. However, in this case the transmission type is not significant. To improve the model, we change the `k` parameter in the `step` function to $log(n)$, where $n$ is the number of observations, so we use the Bayesian information criterion.

This model provides a slightly lower adjusted $R^2$, 0.834, but now all selected independent variables (`wt`, `qsec`, and `am`) are significant. As `wt` and `am` are rather correlated, we decide to include as well an interaction term to the model

```
model.bic2 <- lm(mpg ~ wt + qsec + am + wt:am, data=mtcars)
```

The coefficients are even more significant (see Table 4, Appendix B), and the adjusted $R^2$ value is the highest so far: 0.88 We hence select this last method, which also shows a nice behavior of its residuals (see Appendix C for more details).

## Conclusions

In this report, we have seen that a manual transmission is correlated with a higher fuel economy than automatic transmission. In particular, and considering the transmission type alone, manual transmission increases fuel economy by 7.245 mpg with respect to automatic transmission. This represents a 42.3% increase.

If we account for other confounders, then the effect of the transmission type on fuel economy is much lower, although still relevant. For the best model found without interaction terms, in which aside from the transmission type we also consider the quarter mile time and the weight of the car, a manual transmission increases the fuel economy by 2.936 mpg while keeping the other variables constant. If we include an interaction term between weight and transmission type, then the effect of manual transmission is to increase by 14.079 mpg the fuel economy but correct it with -4.141 mpg for every 1000 lbs.

# Appendix

### A. Correlation matrix and relationship of variables with `mpg`

The following Figure depicts the correlation matrix of all the terms in the `mtcars` dataset. The most correlated variable with `mpg` is `wt`.

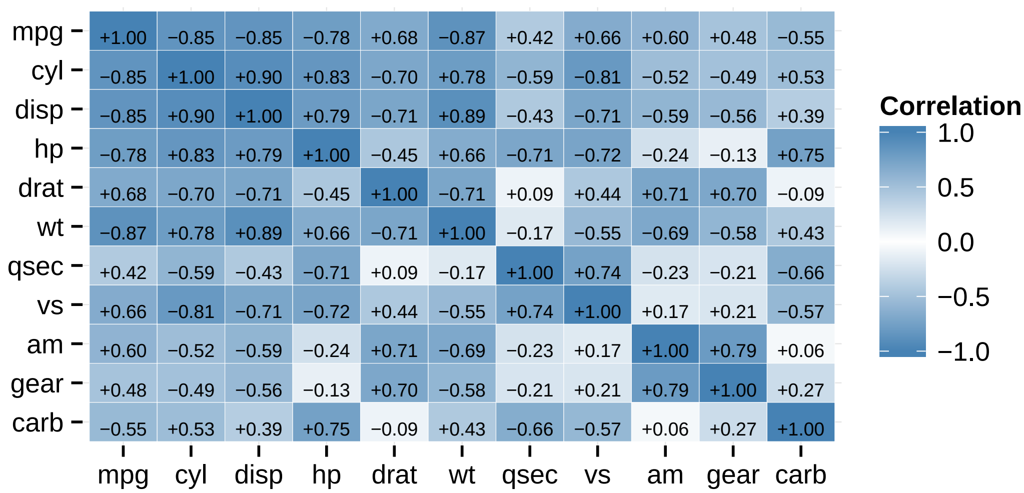


Figure 2: Correlation matrix of the mtcars dataset.

As mentioned in the main text, the following panel shows the relation between fuel economy and the four variables most correlated with it.

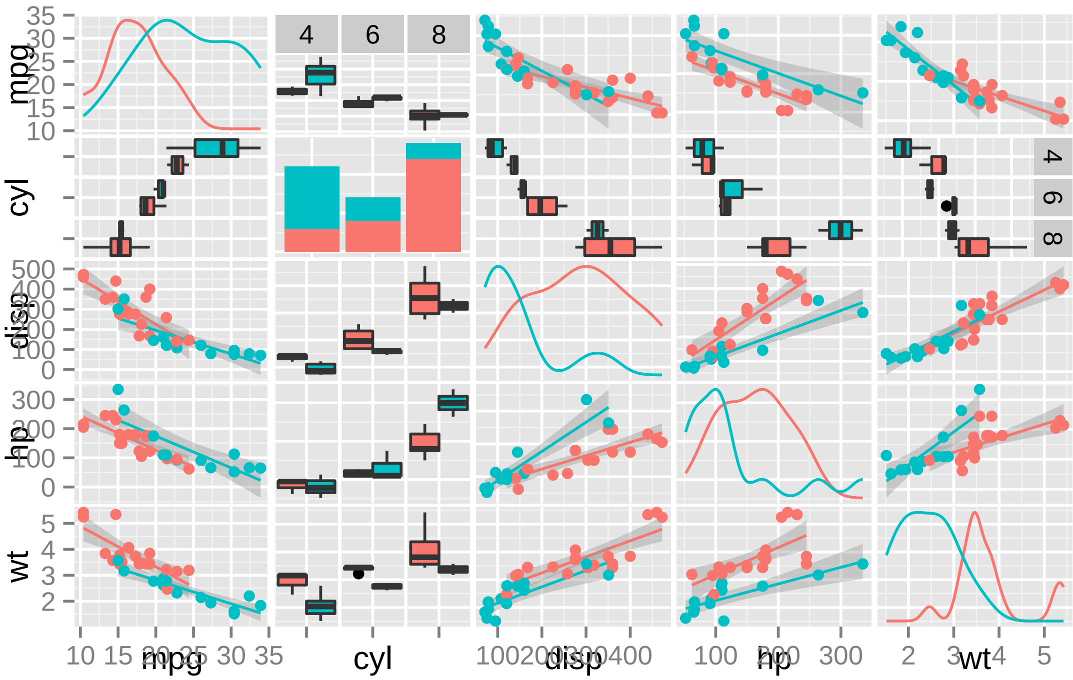


Figure 3: Relationship among several highly correlated variables.

### B. Coefficient tables

The following tables summarise the coefficients of the different models mentioned in the main text.

Table 1: Summary of the coefficients of `model.am`.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| **(Intercept)** | 17.15 | 1.125 | 15.25 | 1.134e-15 |
| **ammanual** | 7.245 | 1.764 | 4.106 | 0.000285 |

Table 2: Summary of the coefficients of `model.aic`.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| **(Intercept)** | 33.71 | 2.605 | 12.94 | 7.733e-13 |
| **cyl6** | -3.031 | 1.407 | -2.154 | 0.04068 |
| **cyl8** | -2.164 | 2.284 | -0.9472 | 0.3523 |
| **hp** | -0.03211 | 0.01369 | -2.345 | 0.02693 |
| **wt** | -2.497 | 0.8856 | -2.819 | 0.009081 |
| **ammanual** | 1.809 | 1.396 | 1.296 | 0.2065 |

Table 3: Summary of the coefficients of `model.bic`.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| **(Intercept)** | 9.618 | 6.96 | 1.382 | 0.1779 |
| **wt** | -3.917 | 0.7112 | -5.507 | 6.953e-06 |
| **qsec** | 1.226 | 0.2887 | 4.247 | 0.0002162 |
| **ammanual** | 2.936 | 1.411 | 2.081 | 0.04672 |

Table 4: Summary of the coefficients of `model.bic2`.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| **(Intercept)** | 9.723 | 5.899 | 1.648 | 0.1109 |
| **wt** | -2.937 | 0.666 | -4.409 | 0.0001489 |
| **qsec** | 1.017 | 0.252 | 4.035 | 0.000403 |
| **ammanual** | 14.08 | 3.435 | 4.099 | 0.0003409 |
| **wt:ammanual** | -4.141 | 1.197 | -3.46 | 0.001809 |

### C. `model.bic2` residuals

The following panel depicts the residuals of the `model.bic2` (see main text for details).

The first plot seems to indicate that the residuals and the fitted values are uncorrelated. The flat distribution of the Scale-Location plot also indicate an homogeneous distribution of the variance of the residuals. The Q-Q plot suggest as somewhat right skewed distribution of the residuals. The final plot indicates that all observations are within a reasonable Cook distance.

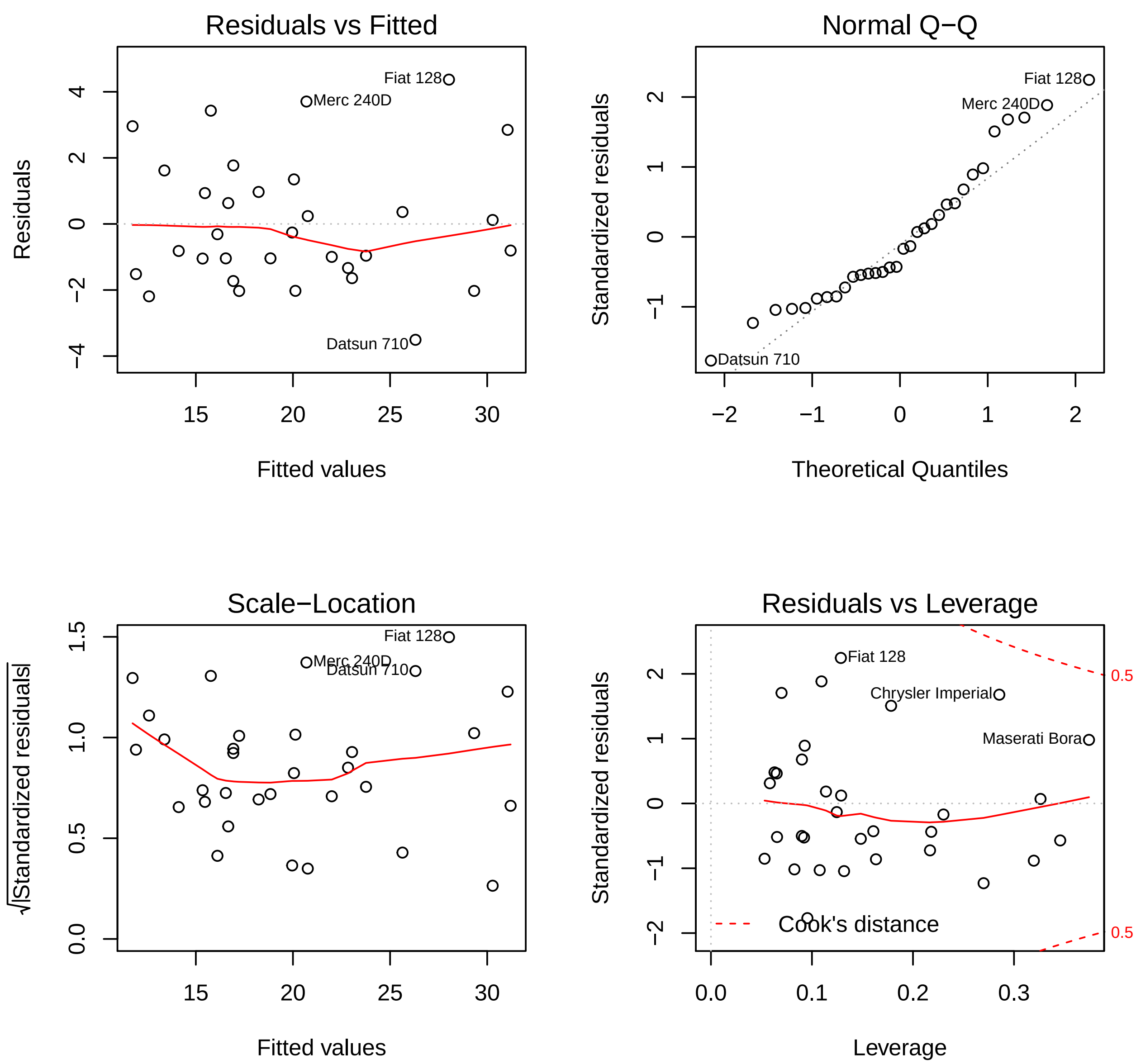


Figure 4: Residuals of `model.bic2`.