# Regression Models Course Project

Maciej Gielnik

18/12/2020

## Executive summary

In this course project I was trying to answer if the type of transmission has an impact on gasoline usage by cars from mtcars dataset. In conclusion the MPG does not depend on transmission type, because 95% confidence interval contain 0. Other variables like weight (wt) and horsepower (hp) with 95% confidence interval have a negative impact on MPG.

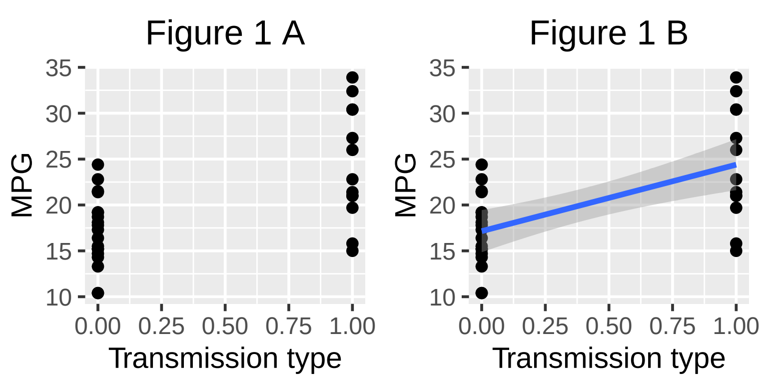
## Introduction

The goal of this report was an attempt to answer the following questions:

- Is an automatic or manual transmission better for MPG?
- Is it possible to quantify the MPG difference between automatic and manual transmissions? If so, how?

## Results

To answer for the questions I have analyzed the mtcars dataset. The type of transmission is located in the mtcars dataset in "am" column. It can take two values: 0 for the automatic transmission and 1 for the manual transmission. In order to inspect if the type of transmission has infuence on MPG I have plotted miles per galon (MPG) in function of transmission using ggplot2 (Figure 1).




The mean of MPG with manual transmission (Figure 1 A, transmission type 1) seems to be much larger than for MPG with automatic transmission (Figure 1 A, transmission type 0). To make more detailed analysis I have performed linear regression (Figure 1 B). The regression line is marked on Figure 1 as blue line and the gray envelope corresponds to 95% confidence interval.

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

```
##              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
```

The intercept corresponds to mean value of MPG with automatic transmission, where the slope corresponds to increase in mean of MPG with manual transmission. The regression suggest that cars with manual transmission can travel on average ~7.24 miles further per gallon of gasoline, however **R^2 test was below 0.36** suggesting a poor model fit.

What other variables can have an impact on gasoline usage? For sure weight of a car - the more car weight, the more gasoline is uses. Also horsepower should contribute to gasoline usage - the more power, the more gasoline car engine needs to perform a mechanical work. Those ideas are supported with exploratory analysis in Figure Supplement 1 A, B. Also preliminary linear regression suggest different gasoline usage in function of horsepower for different transmission types (Figure Supplement 1 C). Similar differences was observed for MPG in function of weight for different types of transmission (Figure Supplement 1 D).

Taking those considerations into account I have created three additional models, where MPG depends on:

- transmission type and horsepower (fit2),
- transmission type, horsepower and weight (fit3),
- transmission type, horsepower, weight and interaction between horsepower and weight (fit4).

The interaction between horsepower and weight has been takes into account because the more car weight, the more power engine needs to speed up the car.

The accuracy of models was tested with the analysis of the variance.

```
anova(fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + hp
## Model 3: mpg ~ factor(am) + hp + wt
## Model 4: mpg ~ factor(am) + hp + wt + I(hp * wt)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 98.963 1.588e-10 ***
## 3     28 180.29  1     65.15 13.560  0.001019 **
## 4     27 129.72  1     50.57 10.526  0.003130 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The residual sum of squared decreased with including horse power, weight and interaction between horsepower and weight, indicating better fit of models to data.

In order to discuss the meaning of coefficients, the 95% confidence interval was calculated for each coefficient.

```
summary(fit4)$coeff
```

```
##                Estimate  Std. Error     t value     Pr(>|t|)
## (Intercept) 49.45224079 5.280730731  9.36465866 5.694894e-10
## factor(am)1  0.12510693 1.333430965  0.09382333 9.259423e-01
## hp          -0.11930318 0.026549992 -4.49352965 1.187315e-04
## wt          -8.10055755 1.789325217 -4.52715777 1.084926e-04
## I(hp * wt)   0.02748826 0.008472529  3.24439879 3.130390e-03
```

```
confint(fit4)
```

```
##                     2.5 %      97.5 %
## (Intercept)  38.61707633 60.28740526
## factor(am)1  -2.61086742  2.86108128
## hp           -0.17377926 -0.06482709
## wt          -11.77194963 -4.42916547
```

```
## I(hp * wt)    0.01010407  0.04487245
```

**In summary:**

- Within 95% confidence interval it is impossible to conclude whether automatic or manual transmission results in MPG, because the confidence interval inculdes 0.
- With 95% confidence interval increase in one horse power decrease the MPG, on average 0.17, holding other coefficients constant,
- With 95% confidence interval increase in one unit of wt (1000 lbs) decrease the MPG, with average 8.1, holding other coefficients constant,
- With 95% confidence interval increase in one unit of interaction bewteen horsepower and weight in 1000 lbs increase the MPG, on average ~0.03, holding other coefficients constant,

### Supplementary data

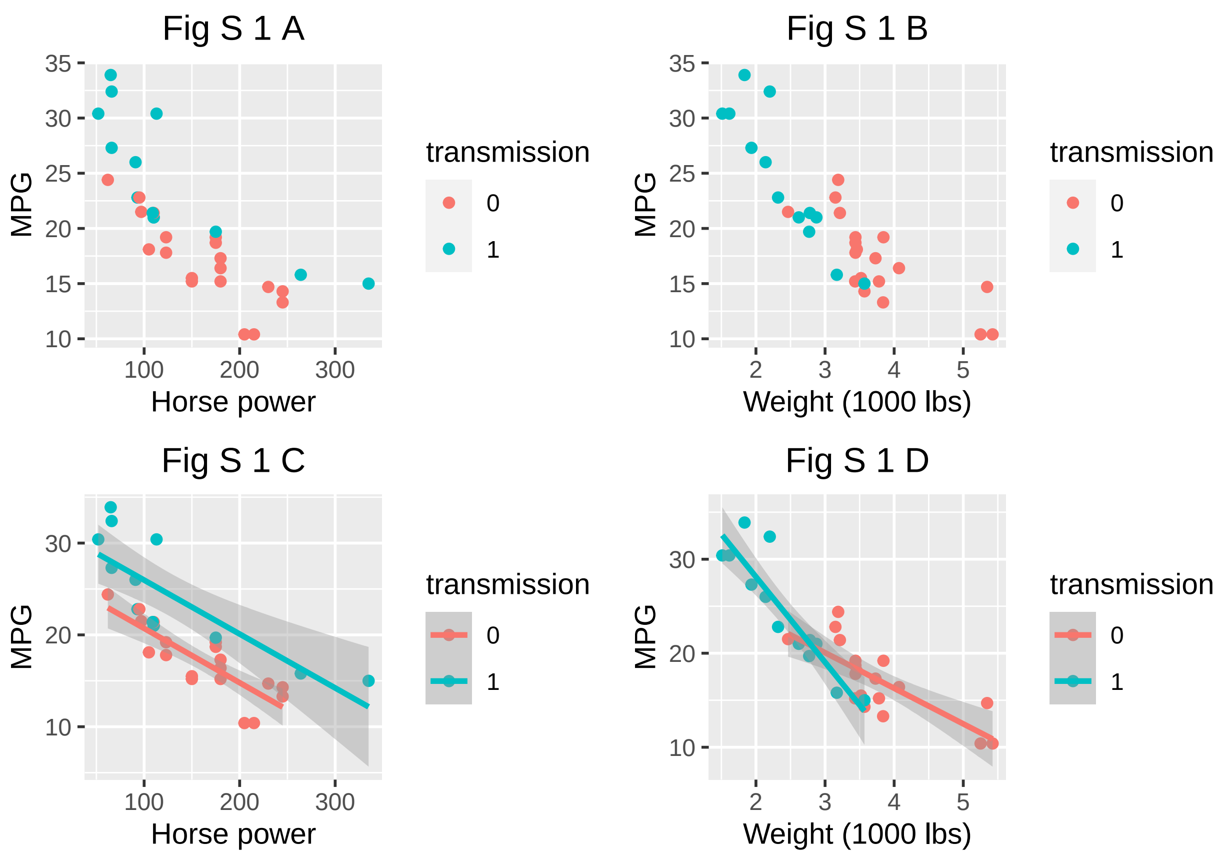


Fig S 2 A − residuals for model 1

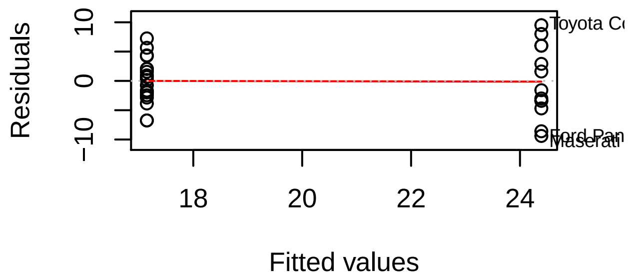



Fig S 2 B − residuals for model 2

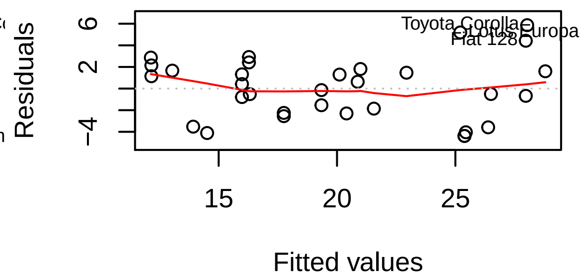



Fig S 2 C − residuals for model 3

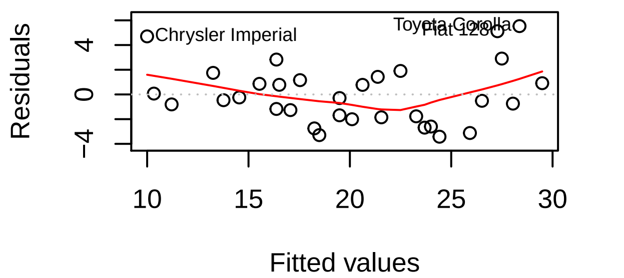



Fig S 2 D − residuals for model 4