# MPG difference between transmissions types

## Executive summary

This report presents results on a problem how transmission type of a car influence on its miles per US gallon (MPG). To explain this relationship, a robust model with transmission type and weight variables is presented. A car with manual transmission has bigger initial MPG (by value of almost 15), but with the increase of its weight, the final MPG decreases almost 3 times more than in case automatic transmission.

## Introduction

This report investigates the relationship between a set of variables and miles per US gallon. In particular, the report focuses on two main issues: 1) is an automatic or manual transmission better for MPG? 2) quantification of the MPG difference between automatic and manual transmissions.

The studied 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)[1]. The data consists of 32 observations on 11 variables: *mpg* - miles/(US) gallon; *cyl* - number of cylinders; *disp* - displacement (cu.in.); *hp* - gross horsepower; *drat* - rear axle ratio; *wt* - weight (lb/1000); *qsec* - 1/4 mile time; *vs* - V/S; *am* - transmission (automatic or manual); *gear* - number of forward gears; *carb* - number of carburetors.

## Exploratory data analysis

A brief look at the data shows that there is a difference between MPG within transmission groups. The unpaired $t$-test for the means unequality in automatic and manual groups has 0.001 $p$-value.

The following table shows values of Pearson correlation between MPG and the other variables, for both types of the transmission and for a given type. The number of cylinders, the displacement, the gross horsepower and the weight are highly correlated to MPG (absolute values greater than 0.75), considered jointly as well as separated. Such variables will be deeply investigated in the next step. A visualization of the relatioship between MPG and the variables is in the **Appendix** section.

Table 1: Pearson correlations between MPG and the other variables (columns) with respect to the types of the transmission (rows).

| | cyl | disp | hp | drat | wt | qsec | vs | gear | carb |
|---|---|---|---|---|---|---|---|---|---|
| all | -0.85 | -0.85 | -0.78 | 0.68 | -0.87 | 0.42 | 0.66 | 0.48 | -0.55 |
| automatic | -0.80 | -0.79 | -0.83 | 0.47 | -0.77 | 0.66 | 0.74 | 0.54 | -0.66 |
| manual | -0.83 | -0.83 | -0.80 | 0.47 | -0.91 | 0.80 | 0.73 | -0.40 | -0.77 |

## Fitting multiple models

The variables mentioned in **Exploratory data analysis** section where selected as potential regressors. Several linear models with combinations of the variables were investigated. Only variants of variable-transsmision type interactions were considered. After the elimination process remained only the models which: have overall $p$-values less than 0.05, each coefficient has also $p$-value less than 0.05, have reasonably high adjusted $R^2$ values (around 0.75 and more) and have arbitrary simple formula. The following models were selected to further evaluation (adjusted $R^2$ values in parentheses):

1. `mpg ~ wt:am` (0.744),
2. `mpg ~ wt:am + am` (0.815),
3. `mpg ~ wt:am + hp:am` (0.805),
4. `mpg ~ disp:am + am` (0.767),
5. `mpg ~ hp:am` (0.735).

[1] Henderson and Velleman (1981), Building multiple regression models interactively. *Biometrics*, **37**, 391–411.

## Final model selection

ANOVA tables were computed for nested models: 1. vs 2., 1. vs 3. and 3. vs 5. All significantly differ. The $p$-values are 0.002, 0.01 and 0.006, respectively.

As the final model was chosen a model with the highest adjusted $R^2$ value, which is model 2. with the 0.815 value.

### Coefficients interpretation

Coefficients of the final model are as follow:

```
(Intercept)       ammanual wt:amautomatic    wt:ammanual
      31.42          14.88          -3.79          -9.08
```

An interpreation of the coefficients is following:

1. we treat 31.42 as the starting MPG of a car without knowledge about its weight and transmission type parameters,
2. we estimate an expected 14.88 increase in MPG in case of manual transmission in holding remaining variables constant,
3. we estimate an expected 3.79 decrease in MPG for every lb/1000 increase in weight of a car with an automatic transmission in holding the remaining variables constant,
4. we estimate an expected 9.08 decrease in MPG for every lb/1000 increase in weight of a car with a manual transmission in holding the remaining variables constant.

### Residual plots and diagnostics

The residual plot (see **Appendix** section) presents relation between predicted values and residuals. As it can be seen, the model is unbiased and homoscedastic. Studentized Breusch-Pagan test also shows that there is no evidence to suspect heteroscedasticity ($p$-value 0.694).

The most influential instance in the dataset is considered. The change in the predicted response when Chrysler Imperial is deleted in fitting the model is 1.03. The change in individual coefficients when the car is deleted in fitting the model is listed below:

```
(Intercept)       ammanual wt:amautomatic    wt:ammanual
     -0.824          0.584          0.930          0.000
```

### Uncertainty quantification

To assess the precision of the predictions, the residual standard error is used. It represents the average distance that the observed values fall from the regression line. In that case it is 2.59, as it is an uncertainty measurement of the outcome MPG.

### Transmission difference quantification

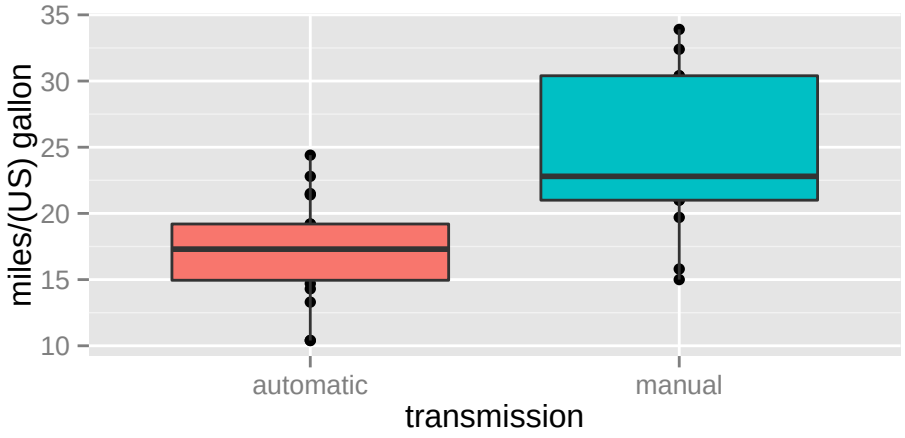
Since it is hard to quantify the difference between transmission types with respect to MPG and other variables, some basic indicators are used. As it was mentioned in **Introduction** section, the difference of MPG averages between transmission groups is statistically significant. The mean MPG for automatic transmission equals to 17.15 and for manual transmission equals to 24.39.

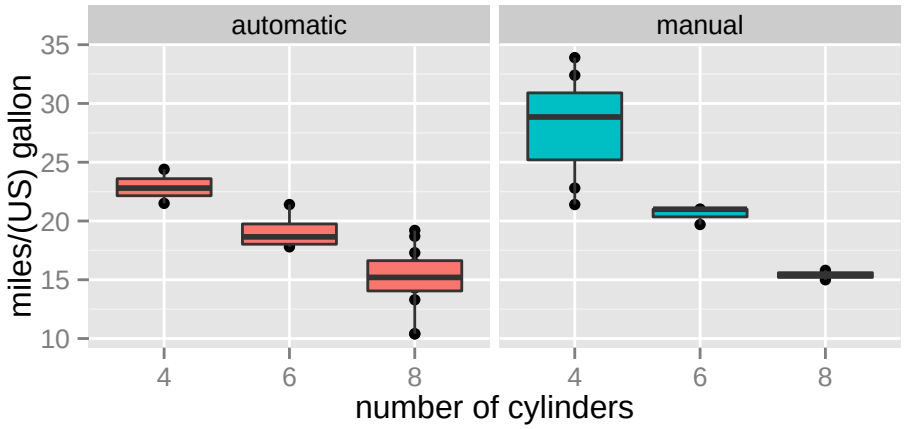
## Summary

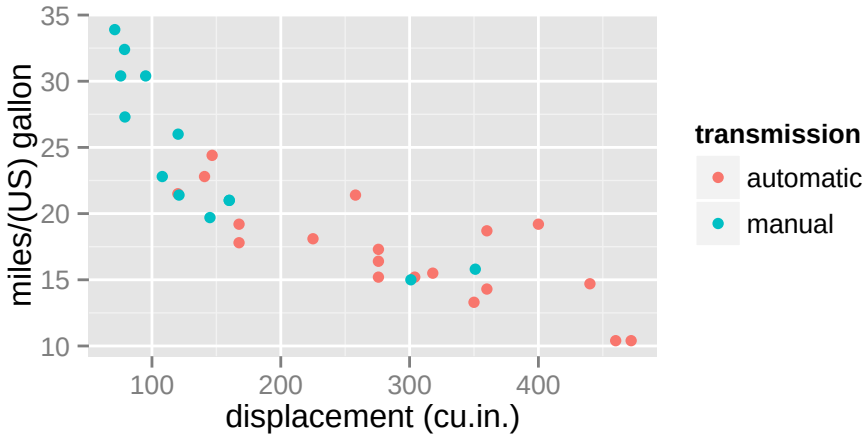
In general, manual transmission is better for MPG. To understand what influences on MPG, the model with transmission type and weight variables was considered. When we investigate the parameters of the cars, that one with manual transmission has bigger initial MPG, but the increase of its weight decreases final MPG almost 3 times more than in case automatic transmission.

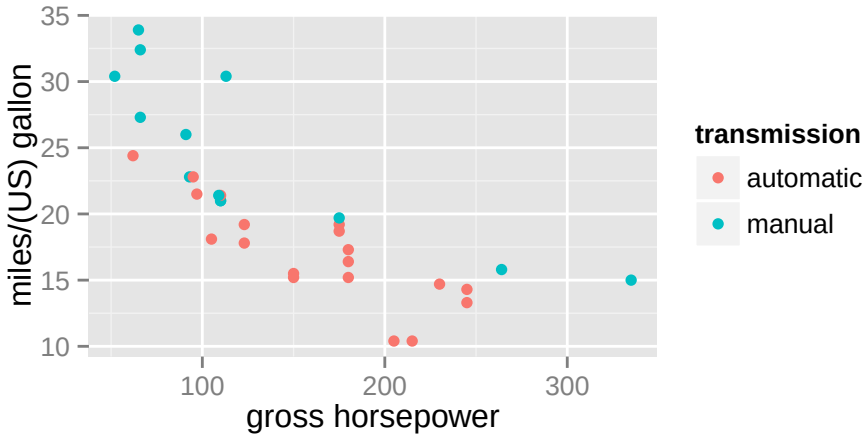
# Appendix

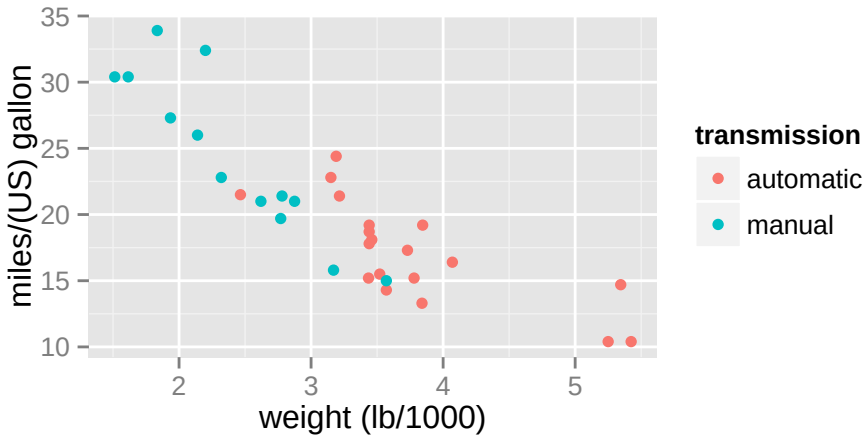
### Exploratory data analysis plots






**Residuals plot**