

# Cars MPG Analysis

*Amade A.*

*December 25, 2015*

## Overview

### Research objective

The objective of this paper is answering the question whether automatic transmission is better for MPG compared to manual transmission. In addition, the researcher also aims to quantify the difference in MPG between cars with different types of transmission.

### Executive summary

This paper analyzes the influence of various car characteristics on fuel economy (measured as miles per gallon of fuel).

## Exploratory Data Analysis

### Dataset

Dataset used for the analysis is **mtcars** dataset, preloaded in R in the *datasets* package. Data originates from the 1974 Motor Trend US magazine, measuring fuel consumption across 10 automobile design aspects for 32 automobiles.

Among the 10 variables, 5 (cyl, vs, am, gear, carb) can be characterized as discrete (factor) variables and remaining 5 (disp, hp, drat, wt, qsec), continuous (numeric) variables. Fuel consumption is represented by the continuous *mpg* variable.

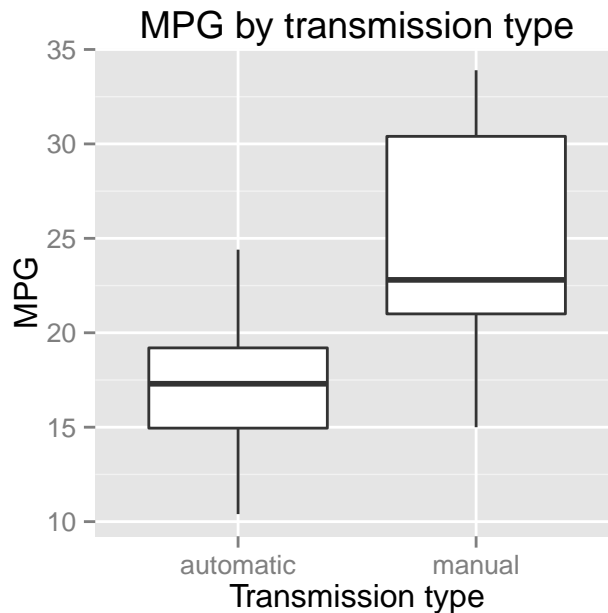
For easier interpretation, variable specifying the transmission type - *am* was recoded, so that the factor levels have straightforward interpretation.

```
library(dplyr, warn.conflicts = FALSE)
mtcars <- mtcars %>%
mutate(am = factor(am, levels = c(0,1), labels = c("automatic", "manual")))
```

### MPG by transmission type

In total, there were 19 cars with automatic transmission and 13 with manual. The following graph shows the differences in measured fuel efficiency among cars with different transmission types.

```
library(ggplot2)
p <- ggplot(mtcars, aes(am, mpg))
p <- p + geom_boxplot()
p <- p + ggtitle("MPG by transmission type") + xlab("Transmission type") + ylab("MPG")
p
```



It can be easily seen, that the fuel efficiency is not the same among the cars with different transmission types, which means that *am* variable is a good candidate for independent regression variable. Median MPG for cars with automatic transmission equals **17.3** and for the ones with manual **22.8**. Values seem to be more concentrated around the median in the case of automatic transmission, compared to cars with manual transmission. Standard deviations of MPG value for two groups equaled respectively **3.83** and **6.17**.

**IF TIME PERMITS, I CAN INCLUDE HERE BOXPLOTS AND SCATTER PLOTS FOR ADDITIONAL VARIABLES**

## Formal modelling

### Simple model (automatic/manual only)

As a first step in modeling a simple linear model with one binary variable *am* was fitted. Parameters of the regression are summarized below.

```
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)$coef
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	17.147368	1.124603	15.247492	1.133983e-15
ammanual	7.244939	1.764422	4.106127	2.850207e-04

Since the model has only one binary variable *am*, the intercept is interpreted as predicted value of MPG for the base level of *am* variable, which in **mtcars** dataset is “automatic transmission”. Value of 17.15 informs us that cars with automatic transmission are expected to reach 17.15 miles for each gallon of fuel. This value is also the average MPG value for cars with automatic transmission.

Change in MPG value for cars with manual transmission is represented by *ammanual* coefficient. Value of 7.24 tells us that the expected (average) MPG for cars with manual transmission is 7.24 higher compared to cars with automatic transmission.

P value for *ammanual* coefficient is very close to 0, informing us that the mean MPG range is significantly different for cars with different transmission type even at 99.9% confidence level.

$R^2$  of the model equaled 0.36, meaning that 36% of the variation in MPG is explained by the regression model.

## **Inclusion of possible confounding variables**

### **Best model selection**

Here perform a likelihood ratio test

## **Uncertainty measures**

### **Analysis of residuals**

Here plot the residuals

### **Confidence intervals for predictor variables**

Specify here the confidence intervals for those variables

## **Conclusions and interpretation**

Interpret the outcome of the model