

# Is an automatic or manual transmission better for MPG?

*Sergey Bushmanov, 08/21/2014*

## Executive summary

To answer the stated question we should go out and collect MPG measures on all brands existing, i.e. total population. However, with the help of regression analysis we can answer this question approximately, with a stated precision, by analyzing a sample. The results obtained by analyzing `mtcars` sample in R show that automobiles with automatic transmission appear to deliver much less MPG. On average, given sample at hand and developed regression model, autos with automatic transmission have  $\bar{x}$  MPG with 95% CI wrapping this average at ... values, and autos with manual transmission have  $\bar{y}$  MPG with 95% CI. Predicted 95% ranges are ... and ... respectively. However, this simple regression model with single predictor is not very efficient in predicting MPG and an alternative linear model is shown.

## Exploratory analysis

Let's explore data visually by plotting MPG for two groups of cars: with automatic and manual transmission.

This cursory visual analysis does suggests that there is difference in MPG due to type of transmission. Let's validate this conjecture with regression analysis.

## Regression analysis 1.

Let's try an ordinary least squares (OLS) regression of MPG ("mpg") on type of transmission ("am"). It should be noted that although MPG is continuous variable, type of transmission is categorical that can take on two states: either "automatic" or "manual". To perform OLS I am treating "am" as dummy variable: 0 for "automatic" and "1" for manual (explicit factorization also possible).

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125   15.247 1.13e-15 ***
## am              7.245      1.764    4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

## Executive summary

1. Boxplot of MPG on two states of am and exploratory data analysis:
  - mean and median, CI, summary by transmission type.
2. Fitting LS linear regression: preliminary diagnostics, interpretation of coefficients and residual plot. ggplot with additional dimensions, `caret::featurePlot`
3. Alternative models. Comparison among models.
4. Uncertainty in the best-fit model.