

# Regression Models Course Project

Milan Gerza

Thursday, September 18, 2014

## Synopsis

Many readers of our automobile industry magazine, the “*Motor Trends*”, are more and more interested in efficiency of fuel consumption. This study explores the relationship between type of transmission (automatic or manual), and MPG (miles per gallon) of a car.

Two main questions will be answered here:

1. “*Is an automatic or manual transmission better for MPG?*”
2. “*Quantify the MPG difference between automatic and manual transmissions*”

## Data analysis

### Data preparation

The dataset is available by name **mtcars**. The format of the data is a data frame with 32 observations on 11 variables. The variables of interest are *mpg* for miles per US gallon and *am* for transmission (0 = automatic, 1 = manual).

| ## |                   | mpg  | cyl | disp | hp  | drat | wt    | qsec  | vs | am | gear | carb |
|----|-------------------|------|-----|------|-----|------|-------|-------|----|----|------|------|
| ## | Mazda RX4         | 21.0 | 6   | 160  | 110 | 3.90 | 2.620 | 16.46 | 0  | 1  | 4    | 4    |
| ## | Mazda RX4 Wag     | 21.0 | 6   | 160  | 110 | 3.90 | 2.875 | 17.02 | 0  | 1  | 4    | 4    |
| ## | Datsun 710        | 22.8 | 4   | 108  | 93  | 3.85 | 2.320 | 18.61 | 1  | 1  | 4    | 1    |
| ## | Hornet 4 Drive    | 21.4 | 6   | 258  | 110 | 3.08 | 3.215 | 19.44 | 1  | 0  | 3    | 1    |
| ## | Hornet Sportabout | 18.7 | 8   | 360  | 175 | 3.15 | 3.440 | 17.02 | 0  | 0  | 3    | 2    |
| ## | Valiant           | 18.1 | 6   | 225  | 105 | 2.76 | 3.460 | 20.22 | 1  | 0  | 3    | 1    |

### Data exploration and interpretation

The quick data analysis shows (see Fig. 1 in Appendix) that there is a clear difference between the automobiles having automatic vs manual transmission. The mean and standard deviation of MPG vs transmission type summarizes following table:

| ## |         | Manual | Automatic |
|----|---------|--------|-----------|
| ## | mpgMean | 24.392 | 17.147    |
| ## | mpgSd   | 6.167  | 3.834     |

### Hypothesis testing

Let us define  $H_0$  as “The difference in means of MPG in automatic vs manual transmission type is equal to 0”. Alternative hypothesis  $H_a$  states that “True difference in means is greater than 0”. The dataset is small and we cannot argue about strict normality of the data, therefore we test using two sample t-test under 0.95 confidence level (ie.  $\alpha=0.05$ ). Since the p-value is less than  $\alpha$  we would reject  $H_0$  in favor of  $H_a$ . Details of the t-test:

```
##
## Welch Two Sample t-test
##
## data: x1 and x2
## t = 3.767, df = 18.33, p-value = 0.0006868
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 3.913 Inf
## sample estimates:
## mean of x mean of y
## 24.39 17.15
```

## Fitting models, diagnostics and interpretation

Both linear and anova models were fitted into the data. The results do not show any significant difference in the model we select, so a linear model was selected. The intercept and predictor values are:

```
## (Intercept)      am
##      17.147      7.245
```

Summary of the linear model:

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15      1.12    15.25  1.1e-15 ***
## am              7.24      1.76     4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36, Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF, p-value: 0.000285
```

The coefficients can be interpreted in a way that change from automatic to manual transmission increases MPG of a car by 7.2449, with signif. level 0.999. As the plot of residual shows (see Fig.2 of Appendix), there is pattern in the variance, showing clear case of heteroscedasticity. This leads to a conclusion that although we fit the linear model, the coefficients will be slightly inaccurate.

## Summary

The analysis shows a probable dependency of MPG on transmission type which we were able to quantify it using linear model. We can therefore answer questions of interest following way: On a significance level 0.95 we have confirmed that manual transmission is better for MPG. The linear model evaluates the difference between automatic and manual transmission is 7.2449 MPG in favor of manual transmission type. However, the real difference may be slightly different due to the unequal variance in groups of observed values.

## Appendix

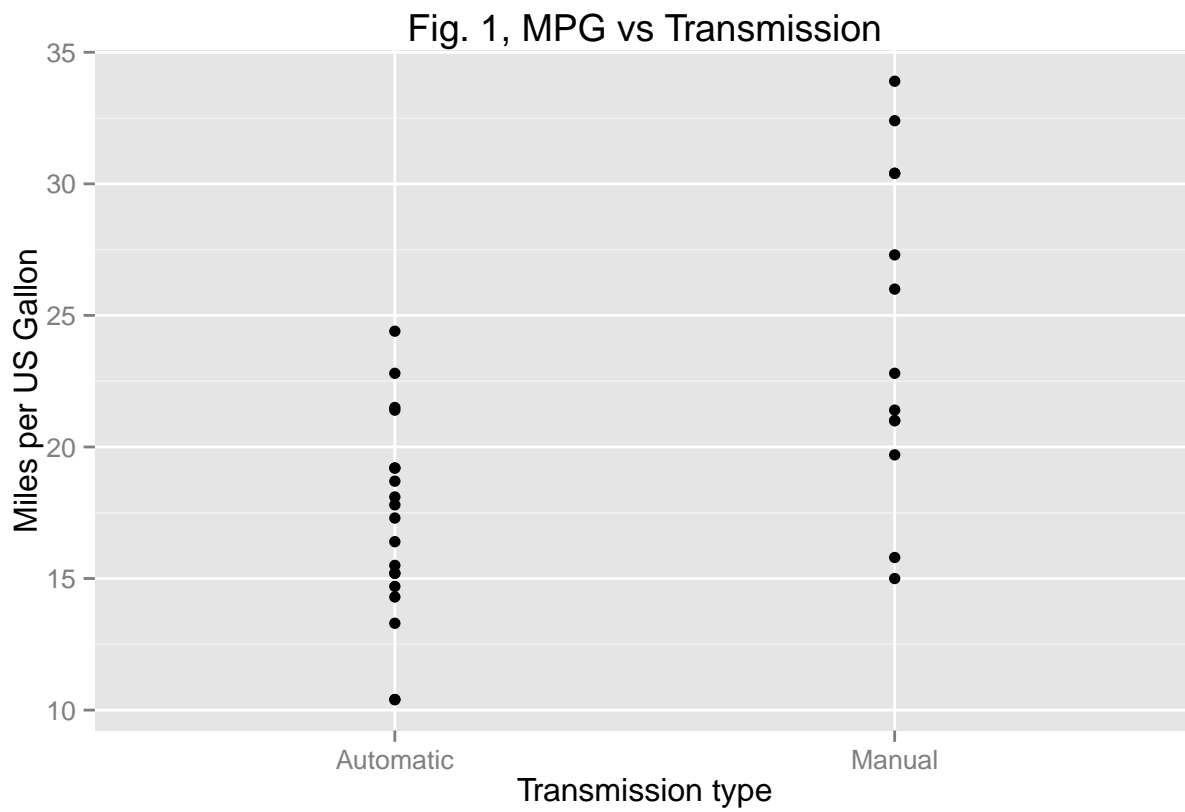


Fig. 2, Residual plot

