Motor Trend analysis - is manual transmission or automatic transmission better for mpg?

Filip Wójcik

Saturday, May 23, 2015

Summary

After performing analysis, the cocnclusion is, that manual transmission results in a bigger mpg value and consequently is a better option. Change from automatic to manual transmission results in a approximately 2.9358 miles/gallon increase (when taken into account along with vehicle weight and calculate on the 1/4 mile time).

Exploratory data analysis

Exploratory data analysis plots (for each variable) can be found as **Figure1** in an appendix.

Is an automatic or manual transmission better for MPG

Figure2 (boxplot of mpg means between two groups) and **figure3 (correlation plot)** from appendix shows the difference in mpg values between two groups - with manual and automatic transmission. From the plot we can conclude that **variances are not equal between the groups**. We can **suppose** (judging the data on the plot by eye) that the manual transmission system comes with higher values of mpg. This leads to two hypotheses:

H0: **mean mpg value is equal** for manual and automatic transmission system

Ha: **mean mpg value is higher** for the manual transmission system

We can perform the following t-test for those:

```
auto.mtcars <- mtcars[mtcars$am == 0, "mpg"]
manual.mtcars <- mtcars[mtcars$am == 1, "mpg"]
t.test(manual.mtcars, auto.mtcars, paired=FALSE, var.equal=FALSE, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  manual.mtcars and auto.mtcars
## t = 3.7671, df = 18.332, p-value = 0.0006868
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.913256      Inf
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

With p-value of 0.0006868 we can safely reject the null hypothesis and assume, that with manual transmission mpg is higher.

Because will all would like to travel more with the same amout of fuel, we can conclude, that **manual transmission** is better for mpg.

Quantify the MPG difference between automatic and manual transmissions

To fill this task, regression analysis will be made. The **slope coefficient for "am"** (transmission type) will describe the difference between two types of transmissions. Let's try a simplest possible model:

```
summary(lm(data=mtcars, formula=mpg ~ am))
```

```
##
## 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
```

P-value of **0.000285** suggests that **am** has significant impact on **mpg** but, this model explains only 35.9% of variance (R-squared, used because with am only model is a single linear regression). More variables (x values) will need to be used, to fill this gap.

Because there are multiple variables, that can (potentially) affect mpg value, a proper model needs to be selected from all possible options. Hovewer, some variables affecting mpg may be **multiplecolinear -** so related one with each other. We have to eliminate them, because they are redundant.

Space limitations for this report does not allow to perform the full analysis step-by-step, but the **algorithm of backwards elimination** proceeds as follows:

1. Find multiple regression coefficients for the full model (all variables)
2. Remove each single variable and re-generate models, quantify residuals and adjusted R^2
3. Generate all possible combinations of 2 variables, iterate through them, and regenerate model, quantify residuals, adjusted R^2 and Akiko Information Criterion (AIC) ... (proceed with combinations of 3 variables, etc. until only single variable remains)
4. From all the models - select the one with the best AIC (best-fitting model)

```
step(object = lm(mtcars$mpg ~ ., data= mtcars), direction = "backward")
```

The final model selected by stepwise, backwards elimination algorithm is:

mtcars$mpg ~ wt + qsec + am

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

This model explains 83.3% of variance and indicates that

Changing from automatic to manual transmission mode changes the mpg by 2.9358

Appendix

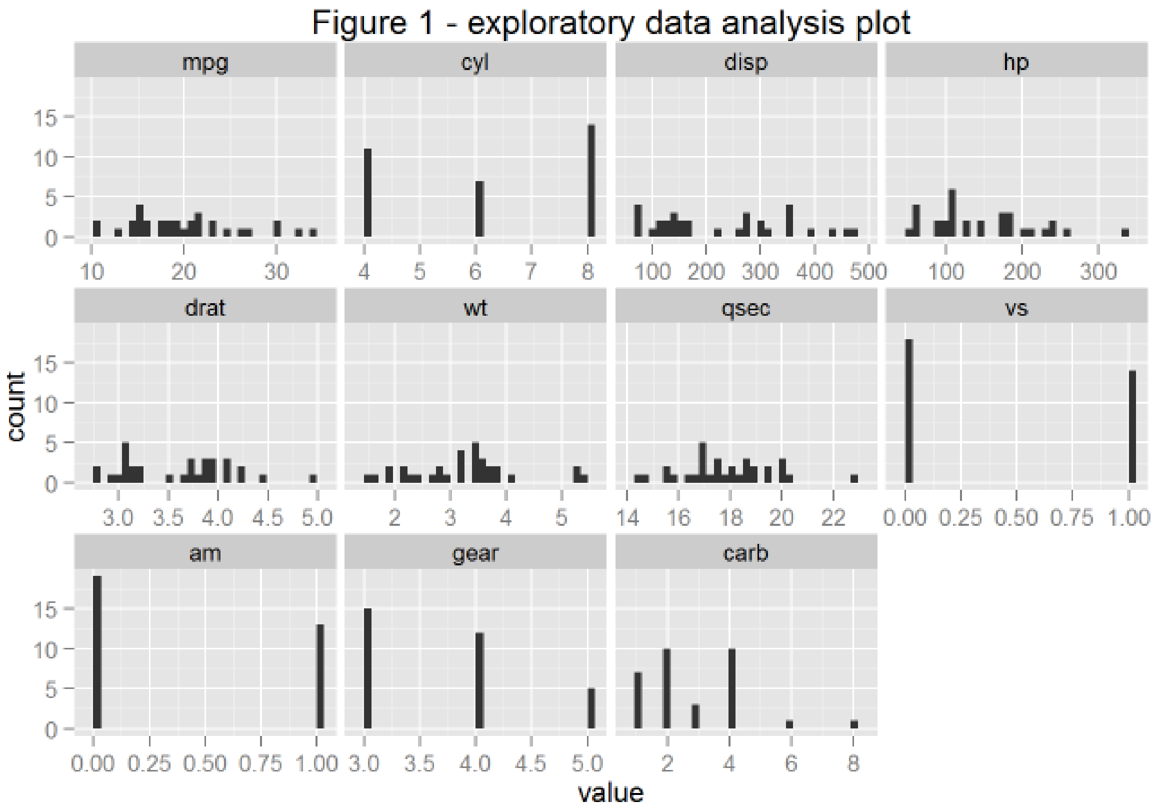
Figure 1 - exploratory data analysis plot
mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
count
value

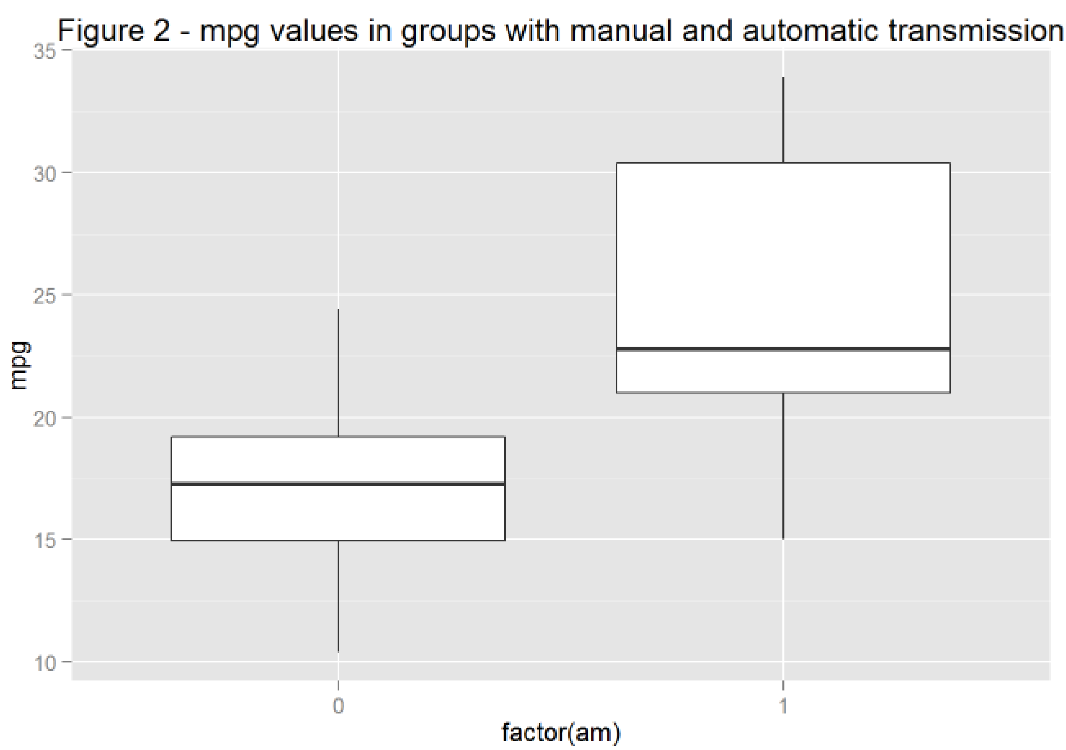
Figure 2 - mpg values in groups with manual and automatic transmission
mpg
factor(am)
0
1

Figure 3 - mpg and am correlation plot

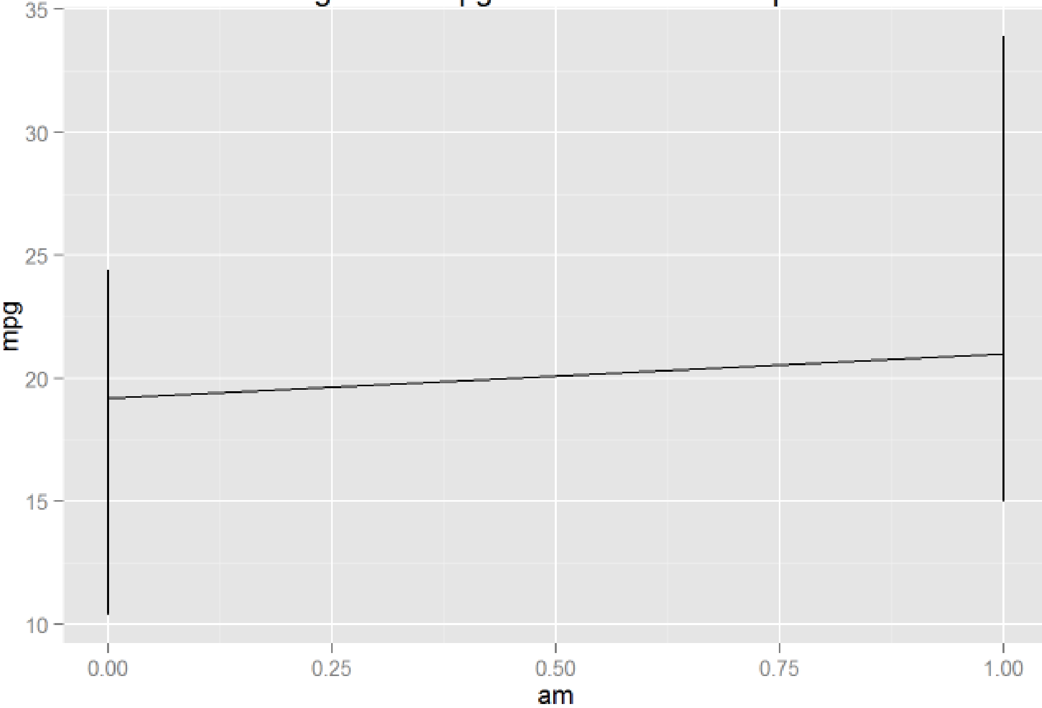

Figure 4-final model plot. Variables correlations