Effect of Transmission Type on MPG of Cars

dogvile

Sep 25, 2015

```
## Executive Summary
```

In this report we will use a dataset from the 1974 Motor Trend US magazine to answer the following questions:

- Is an automatic or manual transmission better for MPG?
- What is the MPG difference between automatic and manual transmissions?

Using data exploratory analysis, hypothesis testing and simple linear regression, we determine whether is a signficant difference between the mean MPG for automatic and manual transmission cars. In brief, based on anova test it will be tested whether a simple, backward, forward regression model is the best fit for our purpose.

Summary of data

The 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). The qualitative variables such as number of cylinders and gears were converted to factors. A description of the variables is available in the appendix.

```
data <- mtcars
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.1.3
```

```
library("ggplot2")
library("knitr")
library("GGally")
```

```
## Warning: package 'GGally' was built under R version 3.1.3
```

```
opts_chunk$set(echo = FALSE)
opts_chunk$set(fig.width = 5)
```

Exploratory analysis

For our analysis it is been used a boxplot in order to display the MPG group by transmision type. In figure 1, it is clear that manual transmission produces more MPG. Next, a pairwise graph (figure 2) was created in order to get a greater intuition of what other variables may be of interest. There is a linear relationship between MPG and each of cyl, disp, hp, drat, wt, qsec, vs, am. It is also used a kind of heatmap of correlation matrix of all variables for better data vizualiation purposes. The covariance was also computed (figure 3) between every variable and the positive values were noted (qsec = 0.419, vs = 0.664, am = 0.600, gear = 0.480). Then a linear model was fit on all the variables to determine which variables should be used in the final models.

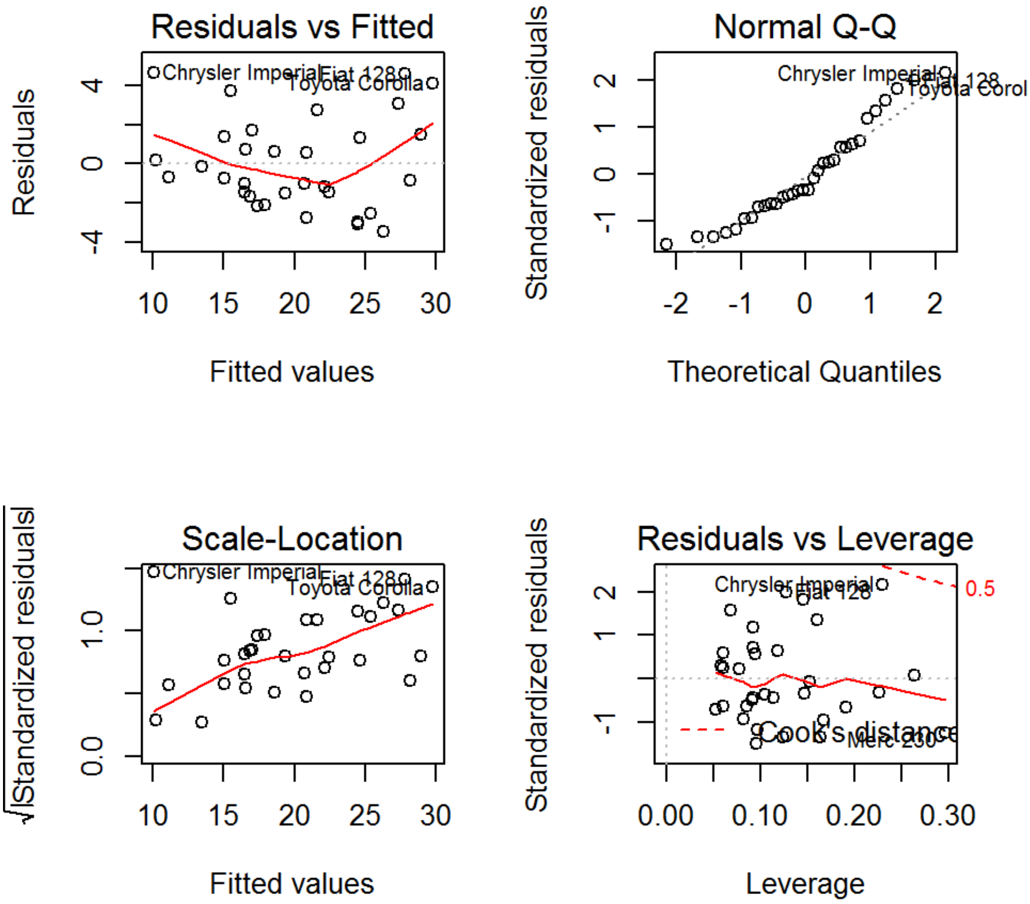
Model

From the initial model, covariance test and visually inspecting the pairwise graph the following variables stood out in particular: qsec, vs, am, wt and gear. Next a stepwise model process was used in order to obtain the most significant predictors to be used. This is done by using the step function which creates multple regression models with different variables and produces list of the best predictors. As shown in figure 5, the most significant predictors in determining the MPG are cyl, hp, wt and am. The summary for this model is show in figure 6, in particular the forumla is given as: lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars). This selection model yielded an R squared value of 84% (figure 6) meaning that very high percentage of variation is explained by the regression model. Next, the new model was compared with a basic model that only uses transmission type as its predictor. A p-value of 1.688e-08 was obtained (figure 7). This value is miniscule which means that the added predictors are significant to improving the model’s accuracy.

```
## [1]        NA        NA 0.8636073
```

Diagnostics

The residuals from the final model are plotted below.



- The Residuals vs Fitted plot shows no pattern between the residuals and fitted values indicating that this regression model is well fit.
- The QQ plot shows that the points line up as expected meaning that the distribtion is normal and our model predictions are accurate.
- In both the Scale-Location plot and the Residuals vs Leverage plots, the points are in a group with none too far from the center indicating no point had too much leverage.

Statistical Inference

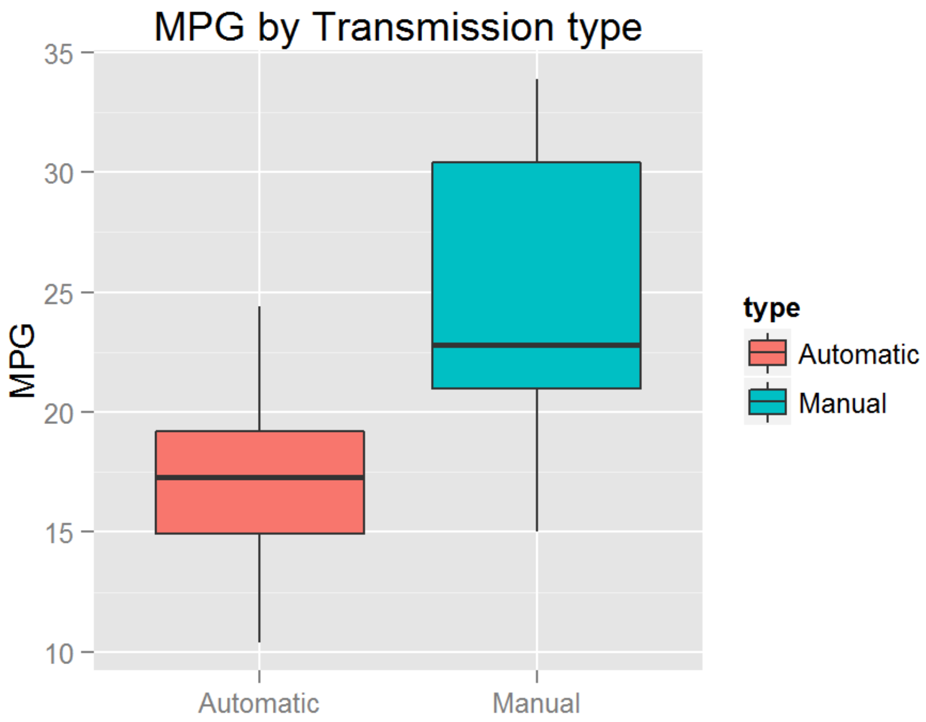
A Two Sample t-test was conducted between the different transmission types. The null hypothesis that transmission types don’t have an effect on the MPG is discarded for a p-value greater than 0.05. The results are shown in figure 8. The p-value of 0.001374 and difference in means show that manual transmission has significantly more MPG than automatic.

Conclusions

The transmission type of a car has a significant effect on its fuel efficiency. According to the model, manual transmission, on average, has 1.81 MPG more than automatics. According to the boxplot, manual transmission has ~ 6 MPG more than automatics.

Appendix

Description of 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 Time to drive per mile
- vs V or ordinary engine
- am Transmission (0 = automatic, 1 = manual)
- gear Number of forward gears
- carb Number of carburetors



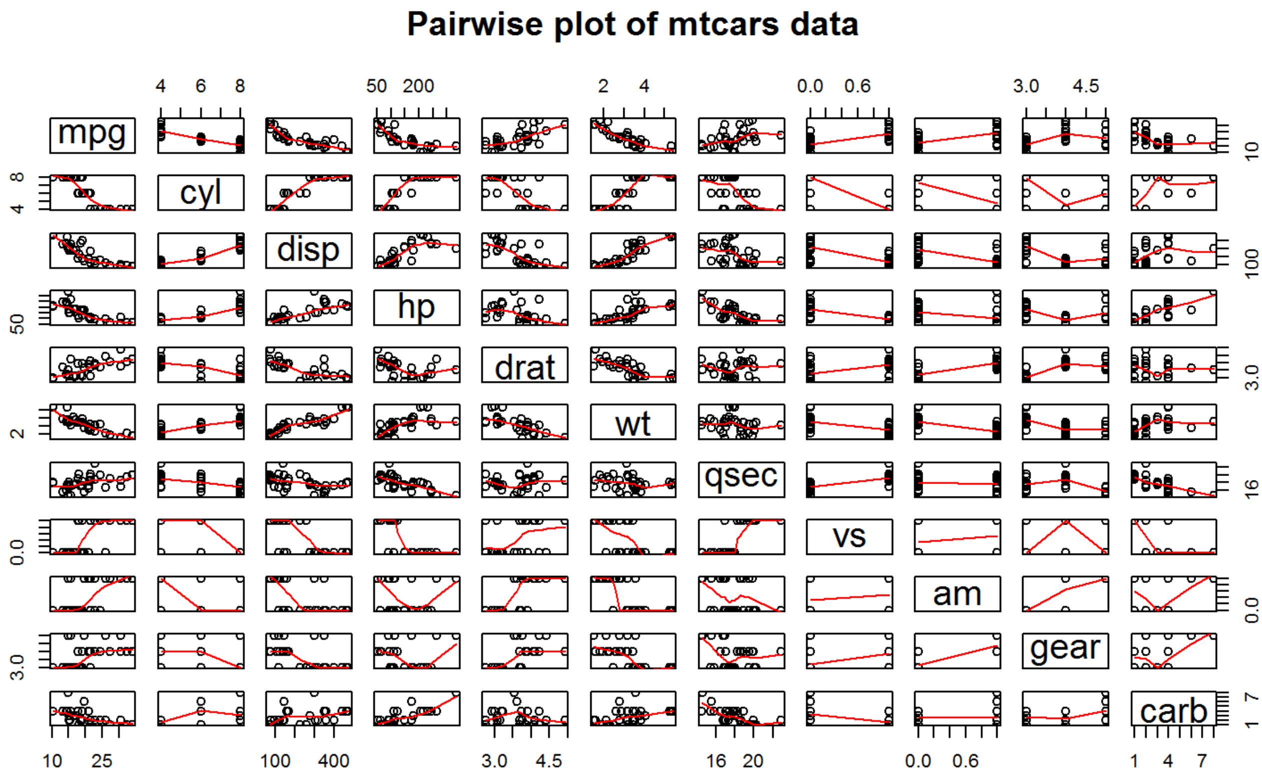


Figure 3

```
head(cov2cor(cov(sapply(mtcars, as.numeric))), 1)
```

```
##     mpg       cyl       disp         hp      drat         wt     qsec
## mpg   1 -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##            vs        am      gear       carb
## mpg 0.6640389 0.5998324 0.4802848 -0.5509251
```

Figure 4

```
everything_model = lm(mpg ~ ., data = mtcars)
everything_model$coeff
```

```
## (Intercept)         cyl        disp          hp        drat          wt
## 12.30337416 -0.11144048  0.01333524 -0.02148212  0.78711097 -3.71530393
##        qsec          vs          am        gear        carb
##  0.82104075  0.31776281  2.52022689  0.65541302 -0.19941925
```

Figure 5

```
new_model <- step(lm(mpg ~ ., data = mtcars), trace = 0)
summary(new_model)$coef
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am           2.935837  1.4109045  2.080819 4.671551e-02
```

Figure 6

```
new_model <- step(lm(mpg ~ ., data = mtcars), trace = 0)
new_model$coeff
```

```
## (Intercept)          wt        qsec          am
##    9.617781   -3.916504    1.225886    2.935837
```

Figure 7

```
basic_model <- lm(mpg ~ am, data = mtcars)
compare <- anova(basic_model, new_model)
compare$Pr
```

```
## [1]           NA 1.550495e-09
```

Figure 8

```
t_test <- t.test(mpg ~ am, data = mtcars)
t_test
```

```
## 
##  Welch Two Sample t-test
## 
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1 
##        17.14737        24.39231
```