Automatic vs. Manual Transmission - Regression Modelling

Robert Jonczy

23.08.2015

Introduction

In this research performed on `mtcars` dataset I will try to answer following questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

Data Transformations

First I load data and perform some data transformations to present some featuers as factors rather than numeric:

```
data(mtcars)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

Data Exploratory

```
str(mtcars[,c('mpg', 'am')])
```

```
## 'data.frame':    32 obs. of  2 variables:
##  $ mpg: num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ am : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
```

```
head(mtcars[,c('mpg', 'am')])
```

```
##                    mpg        am
## Mazda RX4         21.0    Manual
## Mazda RX4 Wag     21.0    Manual
## Datsun 710        22.8    Manual
## Hornet 4 Drive    21.4 Automatic
## Hornet Sportabout 18.7 Automatic
## Valiant           18.1 Automatic
```

```r
cor(datasets::mtcars)["mpg",]
```

```
##        mpg        cyl       disp         hp       drat         wt 
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594 
##       qsec         vs         am       gear       carb 
##  0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```

```r
aggregate(mpg ~ am, data = mtcars, mean)
```

```
##          am      mpg
## 1 Automatic 17.14737
## 2    Manual 24.39231
```

As shown from data manual tranmissions appear to achieve a higher MPG rating as opposed to those of automatic transmissions.

Letting the null hypothesis be: automatic transmissions have a better MPG rating vs manual transmissions. We shall determine if this true by calculating the P-value.

```r
t.test(mtcars$mpg ~ mtcars$am, conf.level=0.95)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  mtcars$mpg by mtcars$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 Automatic    mean in group Manual 
##                17.14737                24.39231
```

The p-value 0.001374 is small (less than 0.05) and confidence interval does not contain 0 shows we have to **reject null hypothesis**. There is significant difference Automatic and Manual transmission, we can say that cars with automatic transmission have lower mpg than manual transmission.

Regression Modelling

First we will try to fit a simple model with `mpg` as outcome and `am` as predictor.

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

```
## 
## 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 ***
## amManual       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
```

Based on the Model only **33.8%** of the variance is explained using the coefficient of determination.

As a next step lets try to fit a model to all variables, and look at the p-values.

```
fit.step <- step( lm(mpg ~ . , data = mtcars), direction = "both", trace = 0, steps=100 )
summary(fit.step)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Based of the model 84 percent of variation of mpg vs tramission type is explained via the coefficient of determination, R2.

R^2 is 0.84 means 84% of the variability explained by this model. Next, we compare linear model mpg ~ am with the best model using ANOVA a.

```
bestfit <- lm(mpg ~ am + wt + qsec, data = mtcars)
anova(fit, bestfit)
```

```
## Analysis of Variance Table
```

```
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(bestfit)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec, 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
## amManual      2.9358     1.4109   2.081 0.046716 *
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ---
## 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
```

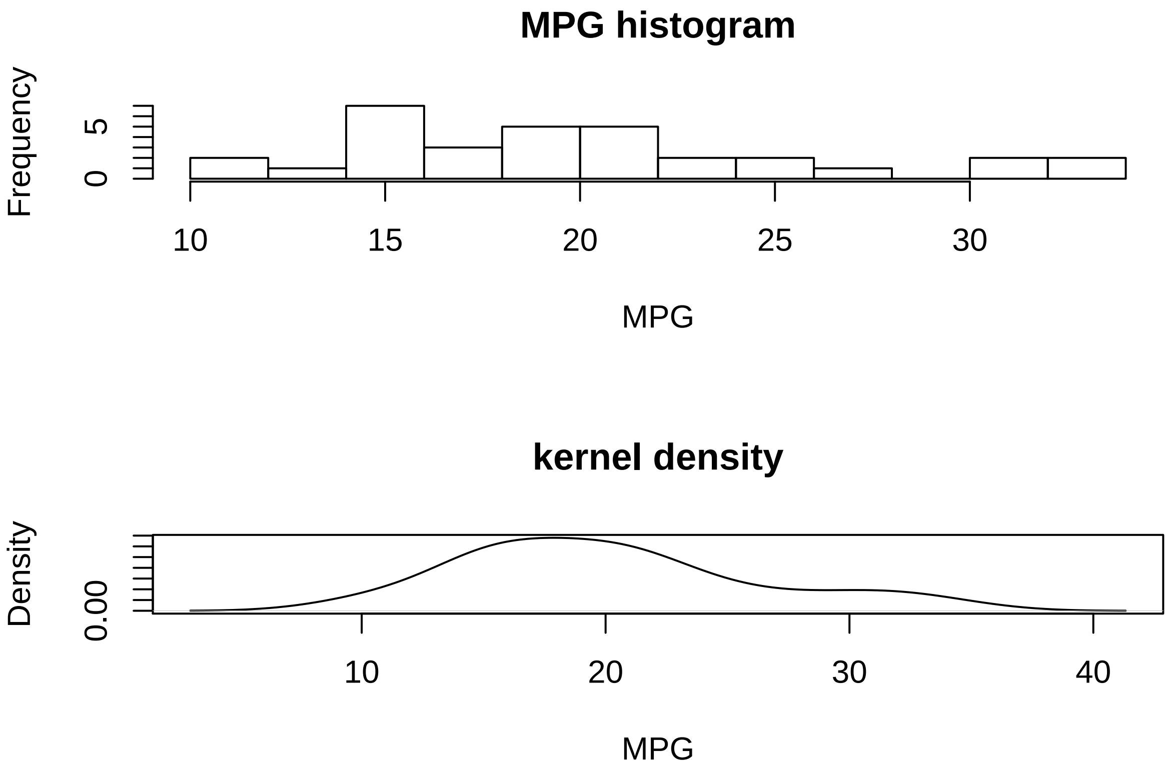
The model explains 84% of difference between mpg and tranmission type; furthermore with an exceptionally low P-value one can conclude that the null hypothesis:Automatic transmissions have a better MPG rating vs Manual transmissions can be rejected, i.e. automatic tranmissions do not achieve a better MPG rating.

Conclusion

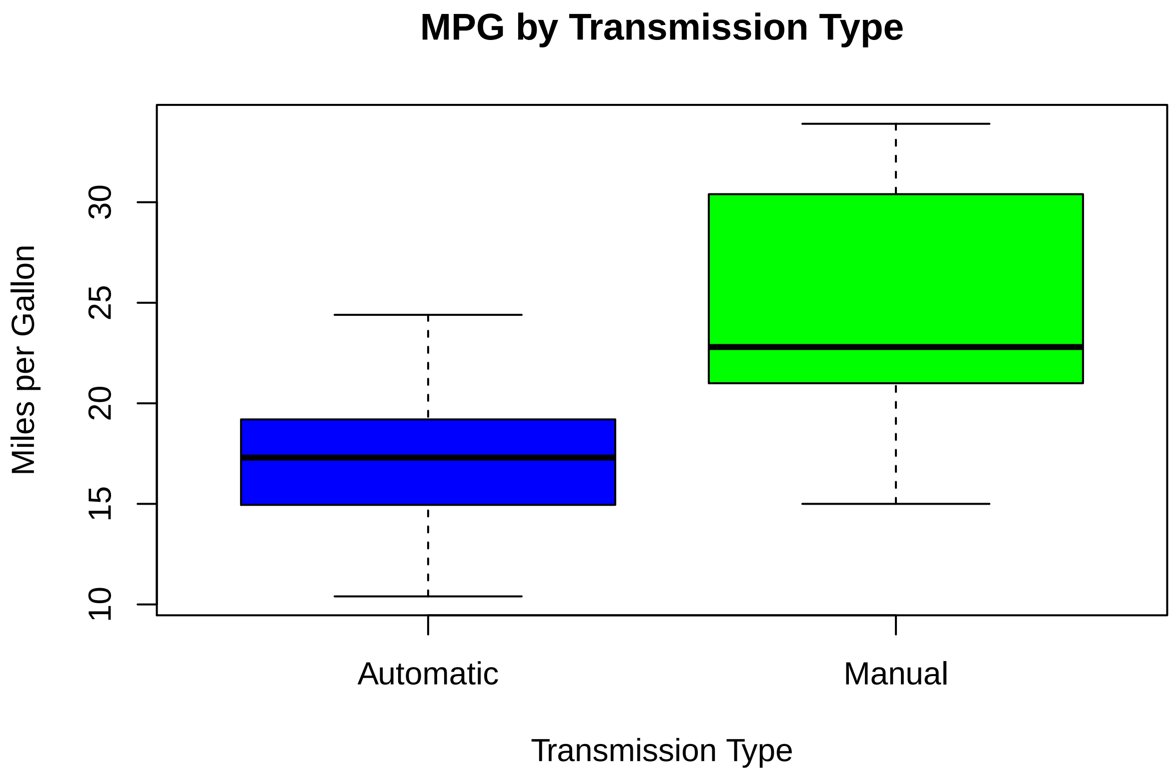
The manual transmision is better than automatic by 1.81. MPG will decrease by 2.49 per 1000 lb. MPG will decrease with bigger number of cylinders

Appendix

Plot1: mpg histogram and kernel density



Plot2: mpg by transmission type



Plot3: residuals