Report on mtcars dataset

Karl Andersson

Friday, April 10, 2015

Summary

This report analysis the dataset "mtcars" in order to answer if an Is an automatic or manual transmission better for MPG and to quantify the MPG difference between automatic and manual transmissions.

We made a regression analysis model that explained 88% of the variance in mpg using the regressors (Wheight, Transmission type, Time on the mile). The result was that manual transmission has a possitive effect for small cars, whereas an automatic gear box has a possitive effect for heavy cars.

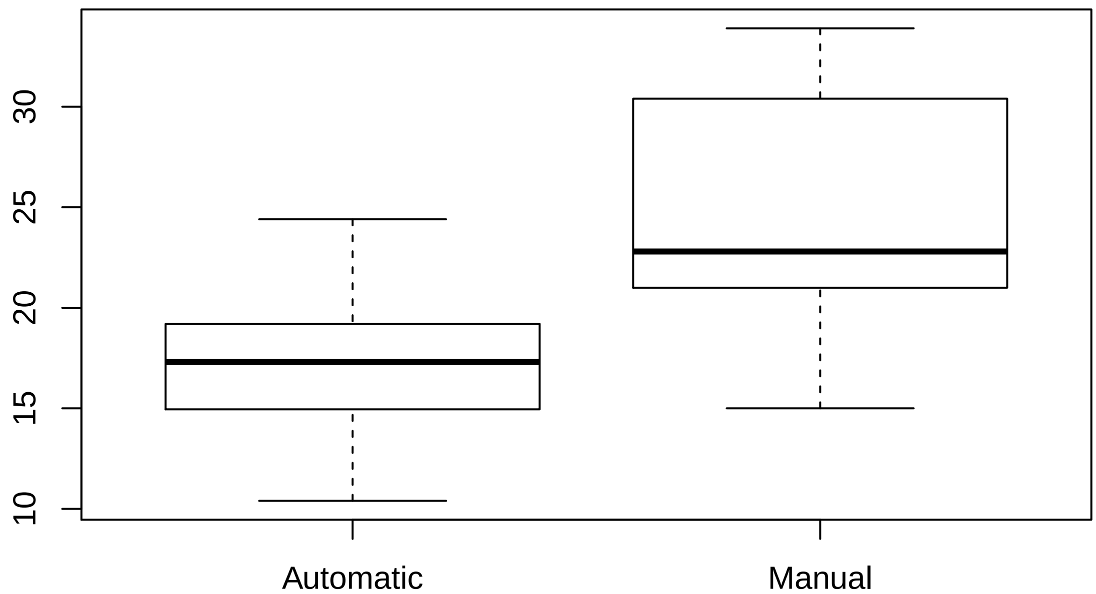
The change in going from Automatic to Manual transmission can be expressed by the formula: Change = 14.08 - 4.14 * Wt. (wt = Wheight of Car[lb/1000])

```r
library(plyr)
library(ggplot2)
data(mtcars)
```

Method and Analysis

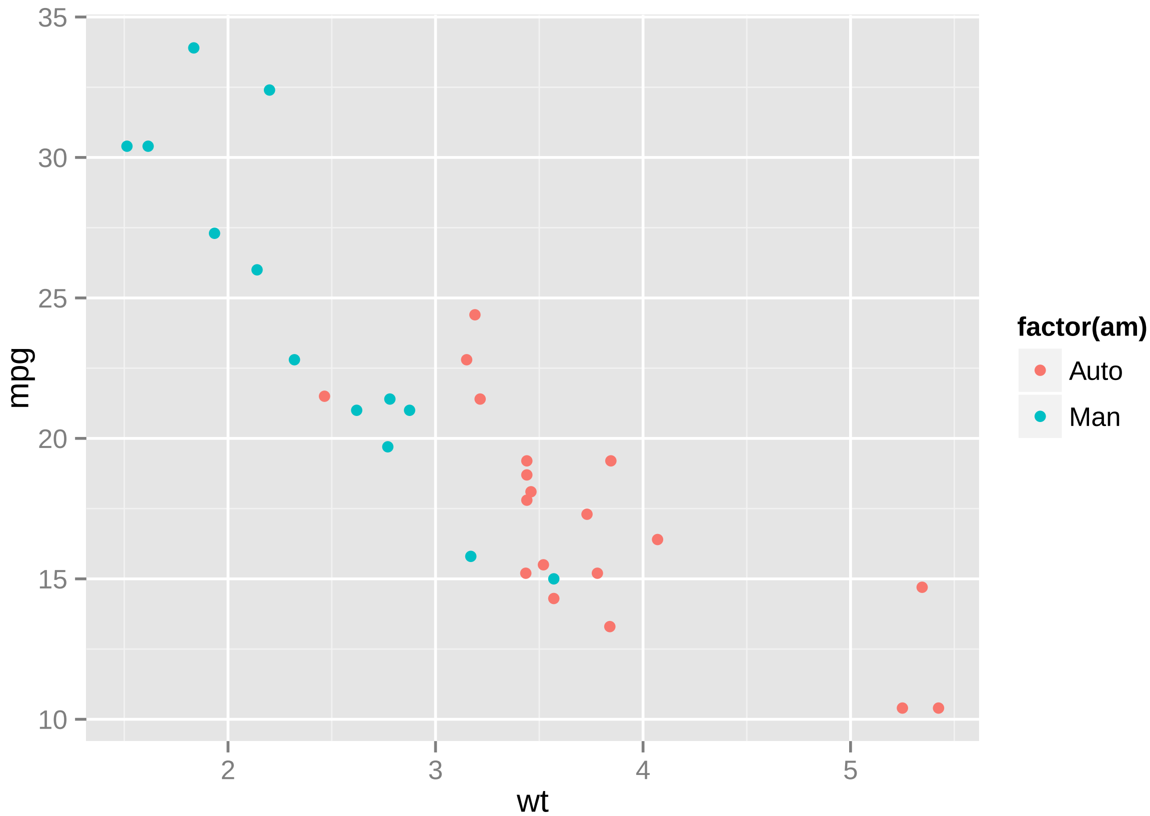
At a first glans it seems as maual gearboxes means higher mpg. Still there can be other factors that are casual and that correlate with what gear box the car has.

```r
mtcars$am = factor(mtcars$am)
mtcars$am = mapvalues(mtcars$am, from = c("0", "1"), to = c("Automatic", "Manual"))
boxplot(mpg~factor(am), data=mtcars)
```



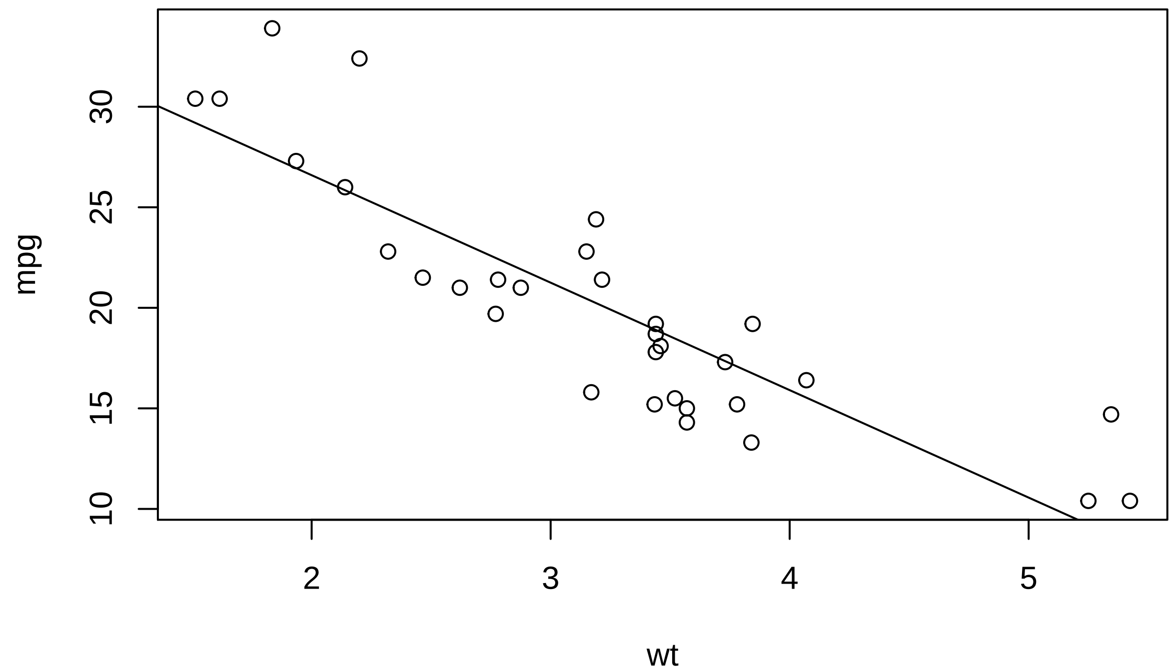
We also se that heavy cars tend to have automatic gearboxes.

```
ggplot(mtcars, aes(x=wt, y=mpg, color=factor(am)))+ geom_point() + scale_colour_discrete(labels=c("Auto
```



We have strong reasons to believe that increase wheight means decreased MPG as F=am (Newton's law of motion) så we need more Force to accelerate a heavier object.

```
fit <- lm(mpg~wt,data=mtcars)
plot(mpg~wt,data=mtcars)
abline(fit$coefficients[1],fit$coefficients[2])
```



Looking att the regression stats we see that using only wheight we explain the mpg variable quite well having a R-Squared value of 0.75 which means that the variable wheight explains 75% of the variance in mpg.

```
summary(fit)$r.squared
```

```
## [1] 0.7528328
```

when makeing a linear regression with 'wt' and 'am' as independents and 'mpg' as outcome 'am' seems quite insignificant. Please note that we factorise the "am" variable.

```
fit = lm(mpg~wt+factor(am),data=mtcars)
summary(fit)$coefficients[,4]
```

```
##     (Intercept)              wt factor(am)Manual
##    5.843477e-13    1.867415e-07    9.879146e-01
```

```
summary(fit)$r.squared
```

```
## [1] 0.7528348
```

We have reason to believe that there is interaction as heavy cars tend to have automatic gear boxes.

```r
fit1 = lm(mpg~wt*factor(am),data=mtcars)
summary(fit1)$coefficients[,4]
```

```
##         (Intercept)                  wt    factor(am)Manual 
##        4.001043e-11        4.551182e-05        1.621034e-03 
## wt:factor(am)Manual 
##        1.017148e-03
```

```r
summary(fit1)$r.squared
```

```
## [1] 0.8330375
```

To get a hint what variables we could have missed we create a model with all available varaibles as regressors and then do a step vise search until we find those significant.

```r
fitFullModel = lm(mpg ~. , data=mtcars)
fitReducedModel = step(fitFullModel, k = log(nrow(mtcars)),trace=F)
```

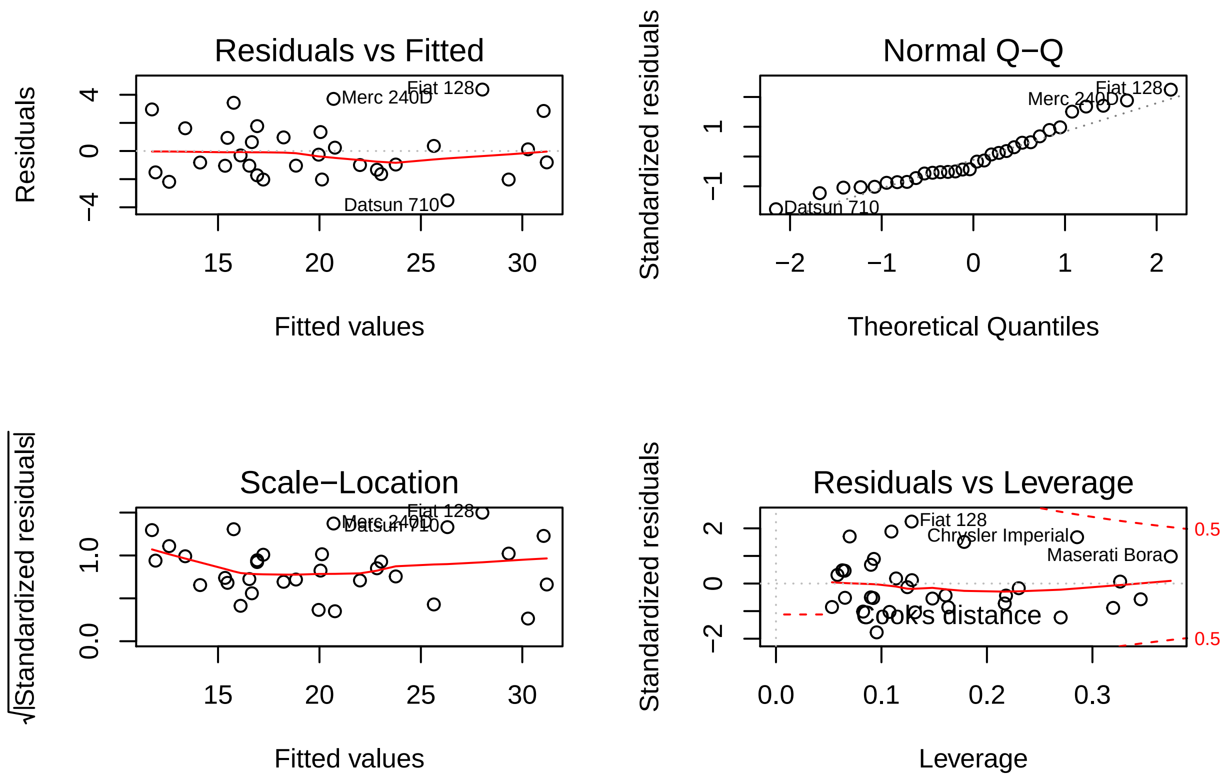
THe output is shoen in Apendix1. We can tell that "qsec" (time to complete a mile) " appears to be a significant varaible, which makes sense because faster cars have stronger engines which usually consume more fuel.Adding Qsec to our original model yields:

```r
fitFinal = lm(mpg~qsec + wt*factor(am), data = mtcars)
summary(fitFinal)
```

```
## 
## Call:
## lm(formula = mpg ~ qsec + wt * factor(am), data = mtcars)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.5076 -1.3801 -0.5588  1.0630  4.3684 
## 
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            9.723      5.899   1.648 0.110893    
## qsec                   1.017      0.252   4.035 0.000403 ***
## wt                    -2.937      0.666  -4.409 0.000149 ***
## factor(am)Manual      14.079      3.435   4.099 0.000341 ***
## wt:factor(am)Manual   -4.141      1.197  -3.460 0.001809 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8804 
## F-statistic: 58.06 on 4 and 27 DF,  p-value: 7.168e-13
```

We can tell that all regressors are significant exeptt for the intercept. Making a residual analysis of the model gives that we don't seem to have any pattern between fit and residuals. And that the residuals are close to beeing normally distributed. Which means that the moduel should be valid.

```r
par(mfrow = c(2,2))
plot(fitFinal)
```



Results

The final model shows that all regressors are signifiacant and the R-Squared Adj is 0.88 which means that the model explains 88 % of variance in MPG. The MPG of a car can be predicted with the following variables and coefficients

```r
fitFinal$coefficients
```

```
##         (Intercept)                 qsec                   wt
##            9.723053             1.016974            -2.936531
##    factor(am)Manual wt:factor(am)Manual
##           14.079428            -4.141376
```

With regards to the impact of the transmission it seems that Manual transmission has a negativ impact on MPG (MPG decreases) for light cars and a Positive impact on heavy cars. THe factor(am)1 coefficient is the change in MPG going from Automatic to Manual but must be considered together with The wt:factor(am) coefficient should be interperated as the change in change of MPG when going from Automatic to Manual that a step in weight give.

Therfore the change in going from Automatic to Manual transmission can be expressed by the formula: Change = 14.08 - 4.14 * Wt. (wt = Wheight of Car[lb/1000])

The change is predicted to be 0 when the car wheight is about 3400 lb. This could be interpreted as the breakingpoint where havier Cars should have Automatic Transmission and lighter Manual Transmission, from a MPG perspective.

Appendix 1

```
summary(fitReducedModel)
```

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

Appendix2

The dataset

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).

A data frame with 32 observations on 11 variables.

- [, 1] mpg Miles/(US) gallon
- [, 2] cyl Number of cylinders
- [, 3] disp Displacement (cu.in.)
- [, 4] hp Gross horsepower
- [, 5] drat Rear axle ratio
- [, 6] wt Weight (lb/1000)
- [, 7] qsec 1/4 mile time
- [, 8] vs V/S
- [, 9] am Transmission (0 = automatic, 1 = manual)
- [,10] gear Number of forward gears
- [,11] carb Number of carburetors