Regression Models Course Project

Executive Summary

Many advancements were made in the design of cars and with it many new features were introduced, but to this day one of the most important aspects of a car buying decision is the MPG value. Although cars nowadays are much more fuel efficient then say 30 yrs ago, the factors that contribute to the mileage per gallon figures are still the same. For example the application of a automatic gearbox on the MPG value might be smaller now than it was 30 years, but it still worth investigating. The model used in this project shows an 2.9 mpg increase when a manual transmission is used. Note that the data used in this project is rather outdated.

Available Data

The data for this project limited to the data available in the mtcars data package. Newer data can be found at an the official U.S. government source for fuel economy. Comparing the fuel efficiency modern automatic gearboxes do come close to the economy of manual gearboxes. This older data set however should make it easier to assess the possibilities of a Linear Model to describe the impact of the gearbox choice on fuel economy.

The dataset contains 32 observations and 11 variables (a detailed description can be found in Appendix 1):

Discussing the Data

Since the data samples and variables small and complete there is no need for any data cleansing. Looking at the data it will become apparent that they are not necessarily independent. As a matter of fact some are highly correlated. As an example the correlation between displacement and number of cyclinders of a car is shown:

```r
cor.test(mtcars$disp,mtcars$cyl)$estimate
```

```
##   cor
## 0.902
```

Other parameters such as hp and wt (weigth) show a high correlation with disp (displacement) for obvious reasons. A heavier car necessitates and stronger engine which in turn requires a larger displacement and hence more cylinders (check for a sample plot in appendix 2).

When we try to quantify the impact of an automatic transmission on the fuel economy we will also need to look at the distribution of manual gearboxes vs the size of the car. The following table shows the amount of cars in percent in each class (4 cylinder, 6 etc) that are equipped with manual transmissions and the mpg means per class(code can be found in the Appendix).

```
##   cyl manual   mpg
## 1   4   72.7 26.66
## 2   6   42.9 19.74
## 3   8   14.3 15.10
```

Creating a Linear Model

We will follow the approach to explore the relationship between mpg and transmission type and use the step function to automatically generate a linear model (lm).

```
lm1 <- step(lm(data=mtcars,mpg~.),direction="backward",test="F", trace=FALSE)

summary(lm1)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.618     6.9596   1.382 1.779e-01
## wt            -3.917     0.7112  -5.507 6.953e-06
## qsec           1.226     0.2887   4.247 2.162e-04
## am             2.936     1.4109   2.081 4.672e-02
```

Weight is indeed has a major impact on the fuel economy of a car (see Appendix 4). Interpreting the results with regards to impact of qsec (as the 1/4 mile time) is a bit more difficult, but overall it can be seen as a compounded variable. Simply stated more hp will result in lower quarter mile times, where the engine output itself is a function of displacement. This, we believe is as good an explanation as any. So a higher quarter mile time will result in a higher MPG, as lower performing cars do you less fuel in general.

Alternate models

We compared many other models with the automatically generated model. But all of them produced p-values for the researched factor than the model above. As an example see the output below where we left the parameter qsec out of the model.

```
lm4 <- lm(data=mtcars,mpg~ wt + factor(am))

summary(lm4)$coef
```

```
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 37.32155     3.0546 12.21799 5.843e-13
## wt          -5.35281     0.7882 -6.79081 1.867e-07
## factor(am)1 -0.02362     1.5456 -0.01528 9.879e-01
```

Using the model produces, one would come to the conclusion that a manual transmission has no impact on the fuel economy, besides the fact that the P-value completely unacceptable.

Results

Following this model we can conclude that the use of a manual transmission has a positive impact on the fuel economy of a car. However, as seen in the model transmission as an indicator has the highest std error (1.4109) of all chosen variables as well as the highest P-value (0.04672), but still below 5% (also check the residual density plot in Appendix 5). According the used model the use of a manual transmission over an automatic will increase the fuel economy by 2.9 mpg. We disregard the results of the second model due to high standard error and the high p-value.

However, there is a problem with the diversity and the limited amount of the data. This does prevent a thorough analysis of the impact of manual transmission. Idealy one would try to keep all variable stable with only the transmission type changing, as would be the case with a side by side comparison of the similar cars, but different transmission types.

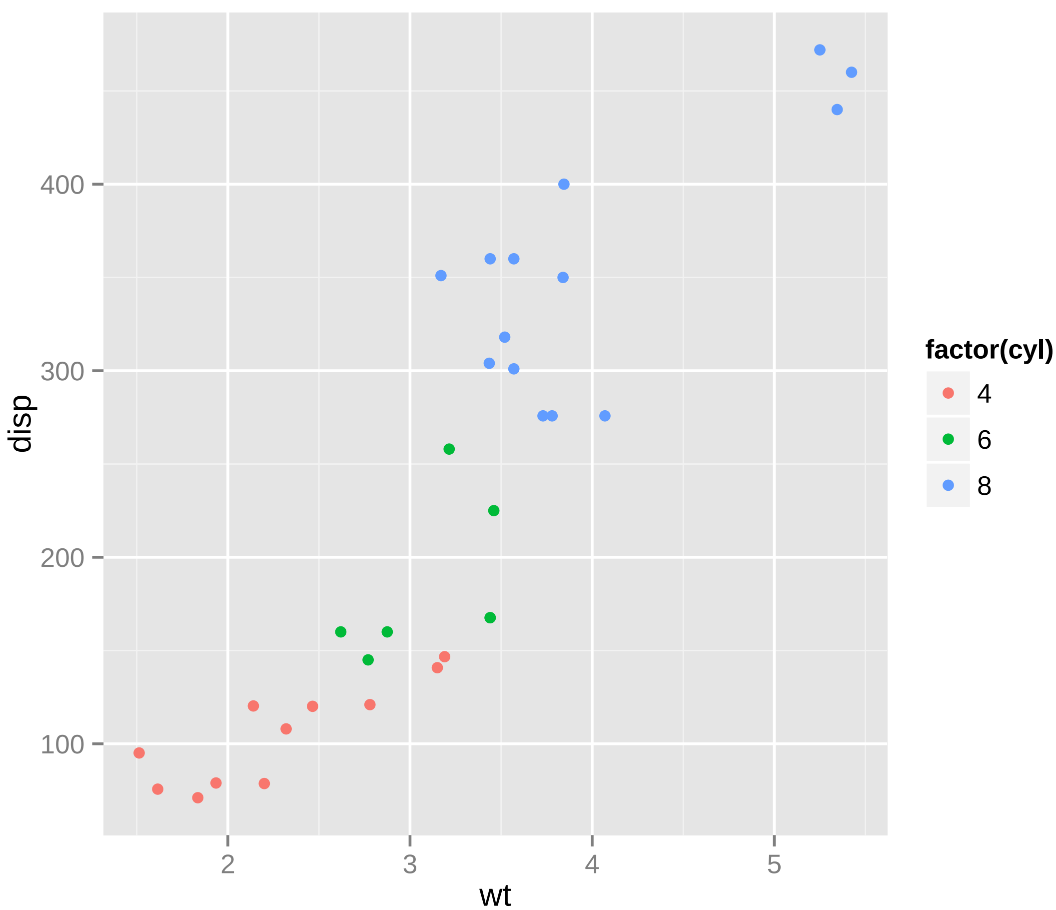
Appendix

Appendix 1

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 carburetor

Appendix 2

```
library(ggplot2)
qplot(wt, disp, color = factor(cyl), data = mtcars)
```



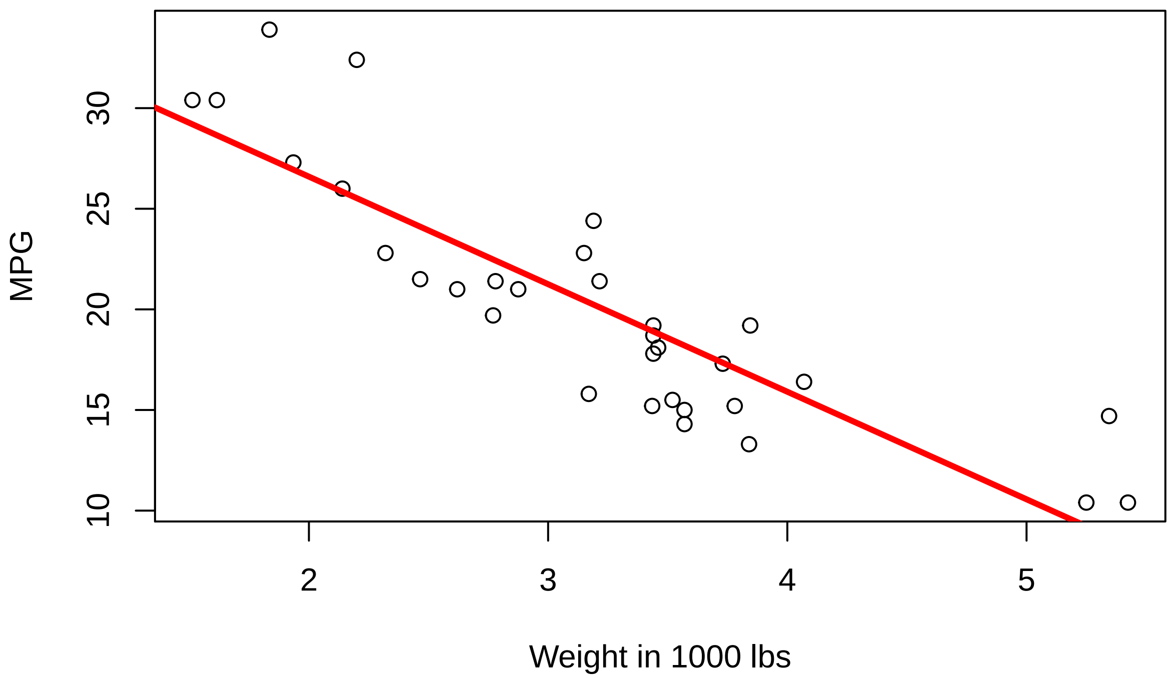
Appendix 3

```r
manual <- c(round(nrow(mtcars[mtcars$cyl == 4 & mtcars$am == 1,])
                  /nrow(mtcars[mtcars$cyl == 4,])*100,1),
round (nrow(mtcars[mtcars$cyl == 6 & mtcars$am == 1,])/nrow(mtcars[mtcars$cyl == 6,])*100,1),
round(nrow(mtcars[mtcars$cyl == 8 & mtcars$am == 1,])/nrow(mtcars[mtcars$cyl == 8,])*100,1))
cyl <- c(4,6,8)
mpg <- c(mean(mtcars[mtcars$cyl == 4,]$mpg),
         mean(mtcars[mtcars$cyl == 6,]$mpg),
         mean(mtcars[mtcars$cyl == 8,]$mpg))
Man_vs_cyl <- data.frame(cyl,manual,mpg)
Man_vs_cyl
```

```
##   cyl manual   mpg
## 1   4   72.7 26.66
## 2   6   42.9 19.74
## 3   8   14.3 15.10
```

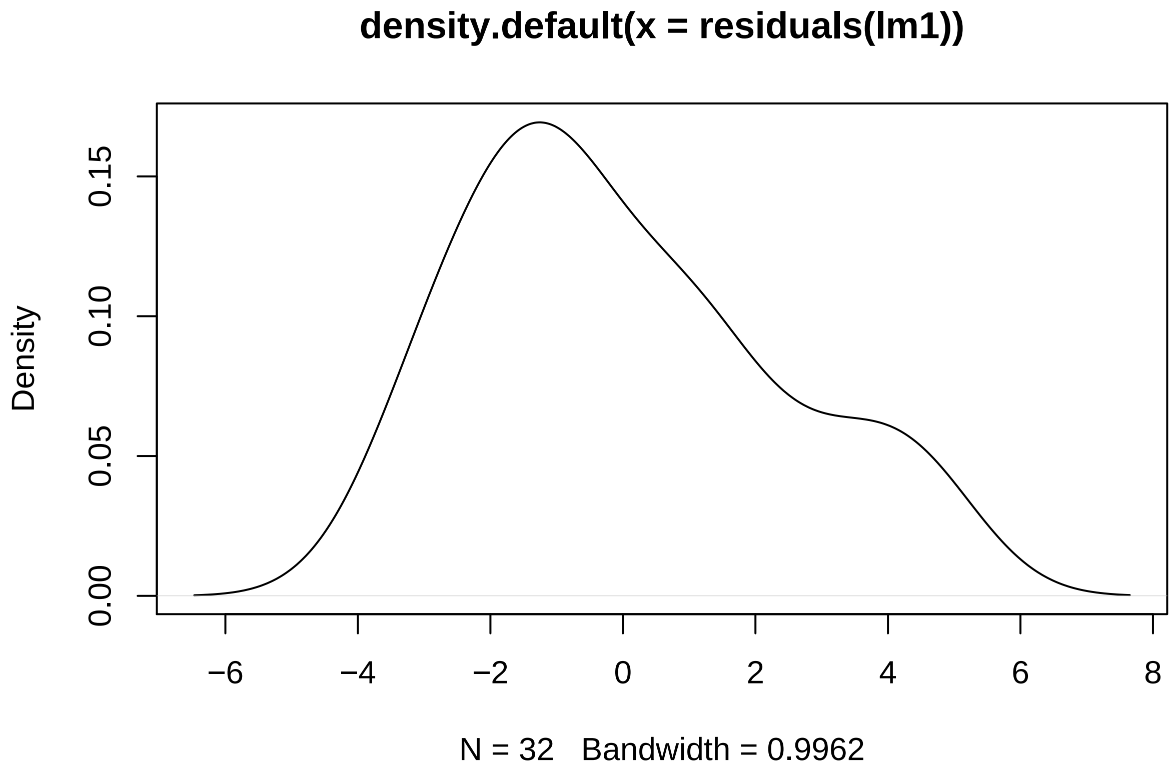
Appendix 4

```r
plot(mtcars$wt, mtcars$mpg, xlab ="Weight in 1000 lbs", ylab = "MPG")
abline(lm(mtcars$mpg~mtcars$wt),col="red",lwd=3)
```



Appendix 5

```r
d <- density(residuals(lm1))
plot(d)
```


density.default(x = residuals(lm1))
Density
0.00
0.05
0.10
0.15
−6
−4
−2
0
2
4
6
8
N = 32 Bandwidth = 0.9962