Regression Models Course Project

Executive summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

1. Is an automatic or manual transmission better for MPG

First, an exploratory data analysis plot is performed. According, to the boxplot shown in Appendix I, the Manual transmission is better for MPG than the Automatic transmissiom type.

Next, a t-test is performed. The null hypothesis is that manual and automatic transmissions have the same performance.

```
library(datasets);
data(mtcars)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
t.test(mpg~am, data=mtcars)
```

```
##
##  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 Automatic    mean in group Manual
##                17.14737                24.39231
```

With a 95 percent confidence interval of [-11.280194 -3.209684] and p-value= 0.001374, we can reject the null hypothesis. The hypothesis testing indicated that a manual transmission is better than an automatic transmission for MPG.

2. Quantify the MPG difference between automatic and manual transmissions

First, a simple linear regression is performed

```
fit1 <- lm(mpg~am, data = mtcars)
summary(fit1)
```

```
##
## 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
```

With R^2 = 0.3598, this model only explains approximately 36% of the variance. Therefore, more regressors need to be considered for explaining the difference between automatic and manual transmission. This, is done using the R function step(), which perform variable selection.

```
fit2 <- step(lm(mpg ~ .,data = mtcars))
```

```
summary(fit2)
```

```
##
## 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
```

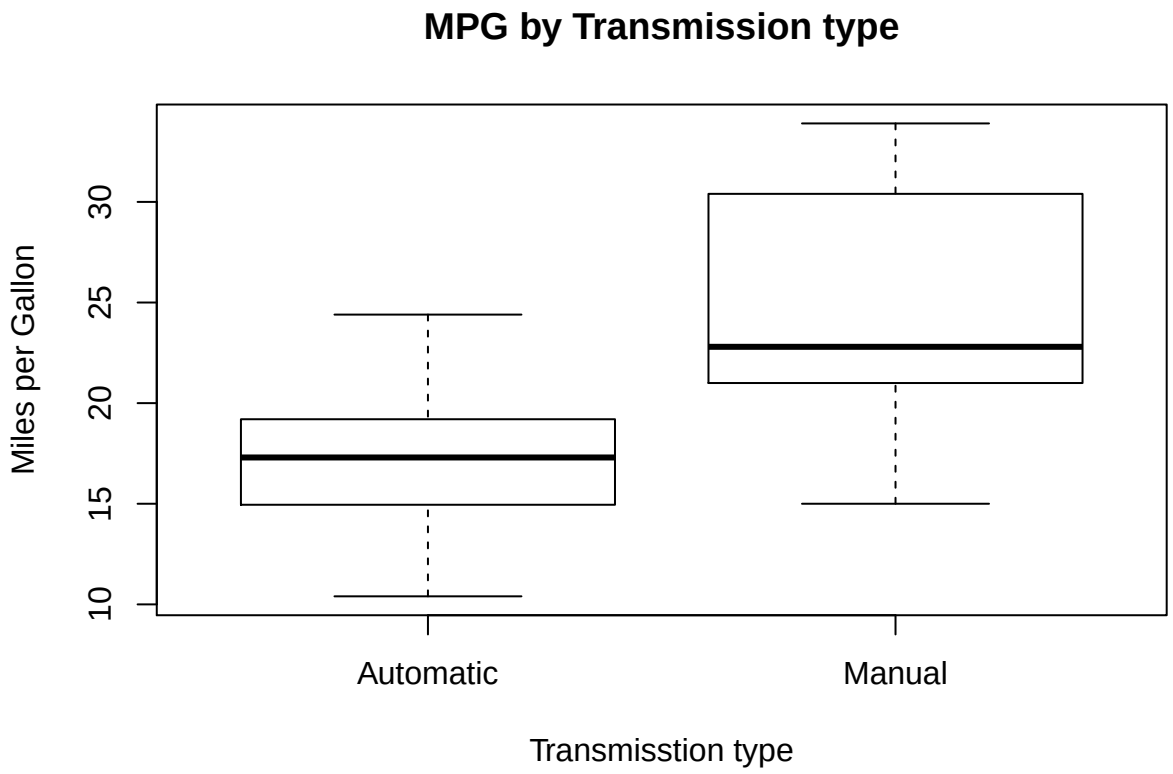
The results show that including am, qsec and wt as regressors 85 % of the variance is explained (R^2 = 0.8497). The best linear regression fit, includes then am, qsec and wt as regressors. Residual plots are shown in Appendix II.

Conclusions

The model which includes three regressors (am, qsec and wt) explains 84% of the variance in MPG The regressor that mostly contribute to mpg, is wt (-3.917), followed by am (2.936) and qsec(1.226). On average, manual transmission cars have 2.936 MPGs more than automatic transmission cars. When only am was included as a regressor the effect was bigger (7.245 MPGs).

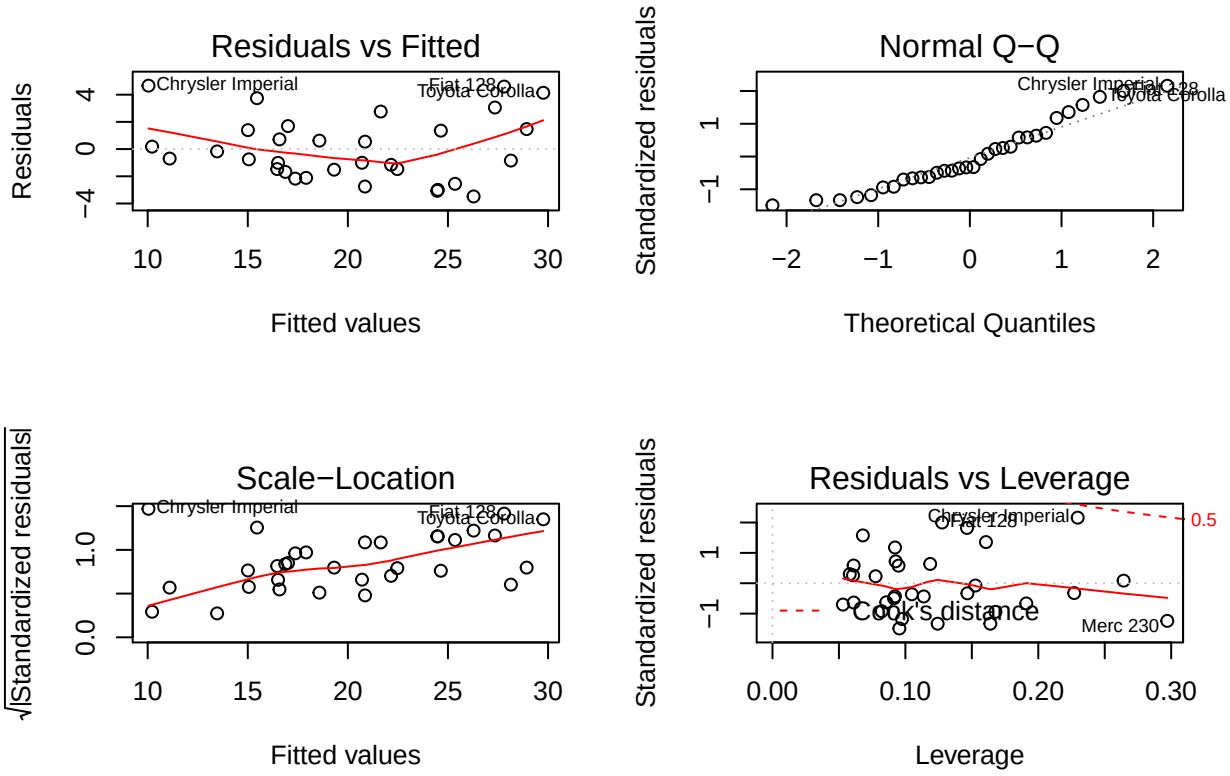
Appendix I

```
boxplot(mpg ~ am, data=mtcars, main="MPG by Transmission type", xlab="Transmisstion type", ylab="Miles p
```



Appendix II

```
par(mfrow = c(2,2))
plot(fit2)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Cook's distance
0.5
Merc 230
Leverage