Regression models on mtcars dataset

Summary

This report is the result of analysis of mtcars dataset made in an effort to answer the following questions: 1. Is an automatic or manual transmission better for MPG? 2. Quantify the MPG difference between automatic and manual transmissions.

The results showed that if we compare mpg only based on the type of transmission, manual transmission is better. However, if we broaden our search of parameters which can influence mpg, we can see, that weight and number of cylinders affect mpg in a more drastic fashion.

One predictor

First we will build a linear regression model based solely on transmission type. Since it is in form of 0s and 1s (representing automatic and manual type respectively), we should consider this as a factor.

```
##             Estimate Std. Error  t value     Pr(>|t|)
## factor(am)0 17.14737   1.124603 15.24749 1.133983e-15
## factor(am)1 24.39231   1.359578 17.94109 1.376283e-17
```

As can be seen from this table, manual transmission (denoted am1) is better than automatic, with difference in mpg = 7.2449393.

More than one predictor

After calculating the correlation between mpg and each of other parameters in the dataset, we can choose those with highest correlation and build a linear regression model on them.

```
##           cyl       disp         hp      drat         wt     qsec
## mpg -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
```

Cyl and disp (number of cylinders and displacement respectively) both showed high correlation with mpg, but since they, obviously, depend on each other, only cyl will be used. Another predictor will be wt.

```
##                Estimate Std. Error    t value     Pr(>|t|)
## factor(cyl)4 33.7535920  2.8134831 11.9970836 2.495549e-12
## factor(cyl)6 29.4962735  3.3130170  8.9031458 1.614393e-09
## factor(cyl)8 27.6744731  3.7983025  7.2860109 7.755263e-08
## wt           -3.1495978  0.9080495 -3.4685309 1.770987e-03
## factor(am)1   0.1501031  1.3002231  0.1154441 9.089474e-01
```

Based on p-value, we can see that if we consider number of cylinders and weight, transmission type can be ignored.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am) - 1
## Model 2: mpg ~ factor(cyl) + wt + factor(am) - 1
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1     30 720.90
## 2     27 182.97  3    537.93 26.46 3.401e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we compare two models using ANOVA, we can notice, that adding number of cylinders and weight as predictors has improved our model.

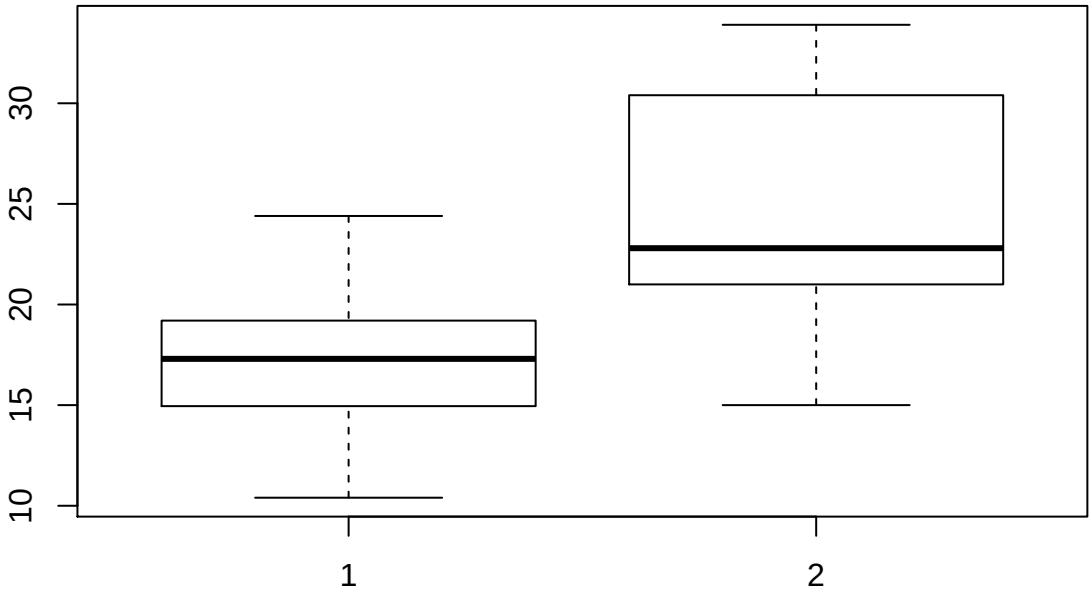
However, if we build our model based on cyl and wt without considering transmission type and compare it to the model, which uses all three parameters as predictors

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(cyl) + wt - 1
## Model 2: mpg ~ factor(cyl) + wt + factor(am) - 1
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 183.06
## 2     27 182.97  1  0.090314 0.0133 0.9089
```

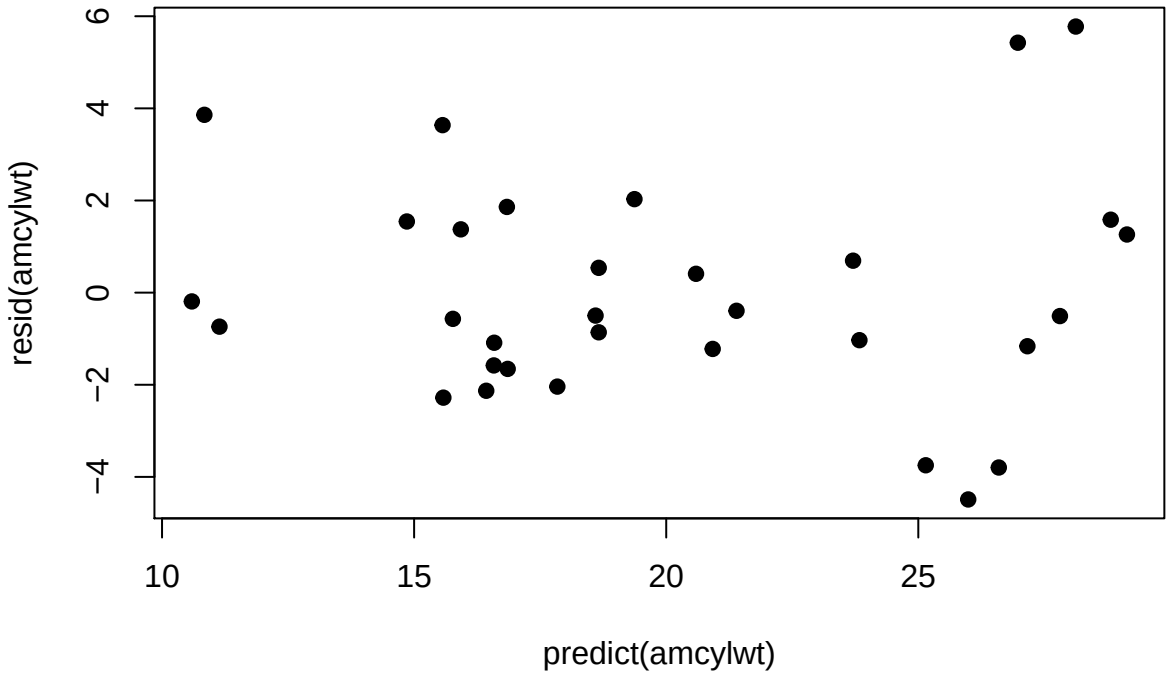
we can see, that our model hasn't really improved.

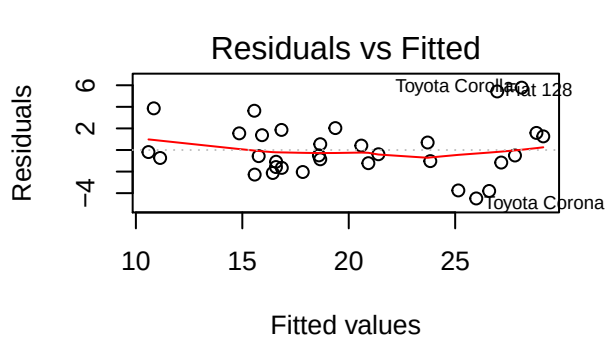
Appendix

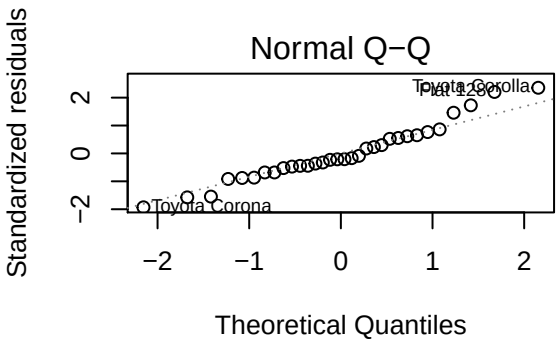
Comparing mpg with automatic and manual transmission type

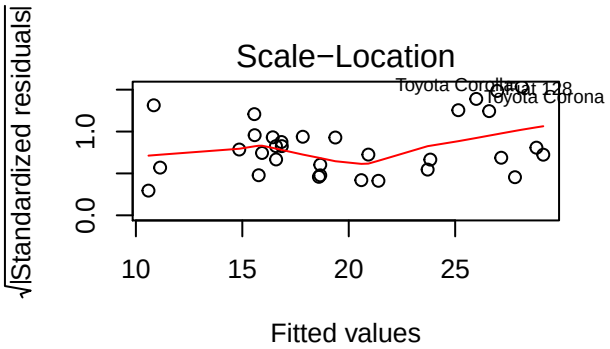


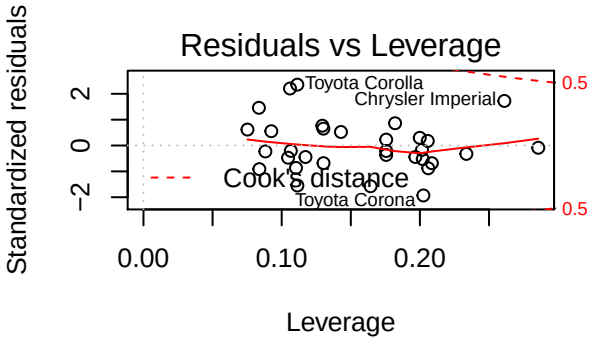
While our p-values showed significance, we can plot a residual plot to check whether there is really no hidden pattern




Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Toyota Corona


Scale−Location
√|Standardized residuals|
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona


Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Cook's distance
Toyota Corona
0.5
0.5