# RegAnalysis_Proj.Rmd

Below we analyze the data for mtcars in R. The goal is to address the following quesion: "Is an automatic or manual transmission better for MPG", and "quantify the MPG difference between automatic and manual transmissions".

## Executive summary

We conducted regression analysis for data mtcars, in order to investigate the significance of transmission type on the mpg. Although an exploratory analysis shows a mean of 17.15 and 24.39 for cars with auto and manual transmission, respectively, a full model including all significant predictors show that the difference in mpg is not significant (1.8, t test not passed with t value 0.2).

## Exploratory analysis

We first investigate the difference in MPG for auto vs manual transmission without including the other variables. The result, analysis below shows different mpg means of 17.15 and 24.39 for cars with auto and manual transmission, respectively (plot attached to end of document).

```
summary(mtcars[mtcars$am == 0,]$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

```
summary(mtcars[mtcars$am == 1,]$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

```
#boxplot(mpg ~ am, data = mtcars, xlab = "Transmission", ylab = "MPG", main="MPG vs Trans Type")
```

## Multiple regression analysis

To further investigate whether the above difference is resulted from other variables, we give a multiple regression to account for all variables. Based also on the Akaike's information criterion, the final model selected is a regression on the following variables (anova output hidden due to lengthiness): cyl, am, hp and wt. The result of the final model fitting is shown (plot attached in the appendix).

```
fit1<-glm(mpg~as.factor(cyl)+vs+am+as.factor(gear)+as.factor(carb)+disp+hp+drat+wt +qsec,data=mtcars)
library(MASS)
#step <- stepAIC(fit1, direction="both")
#step$anova (anova result hidden due to lengthiness)
fit2<-glm(mpg ~ as.factor(cyl) + as.factor(am) + hp + wt - 1, data=mtcars)
#par(mfrow=c(2, 2))
#plot(fit2)
summary(fit2)
```

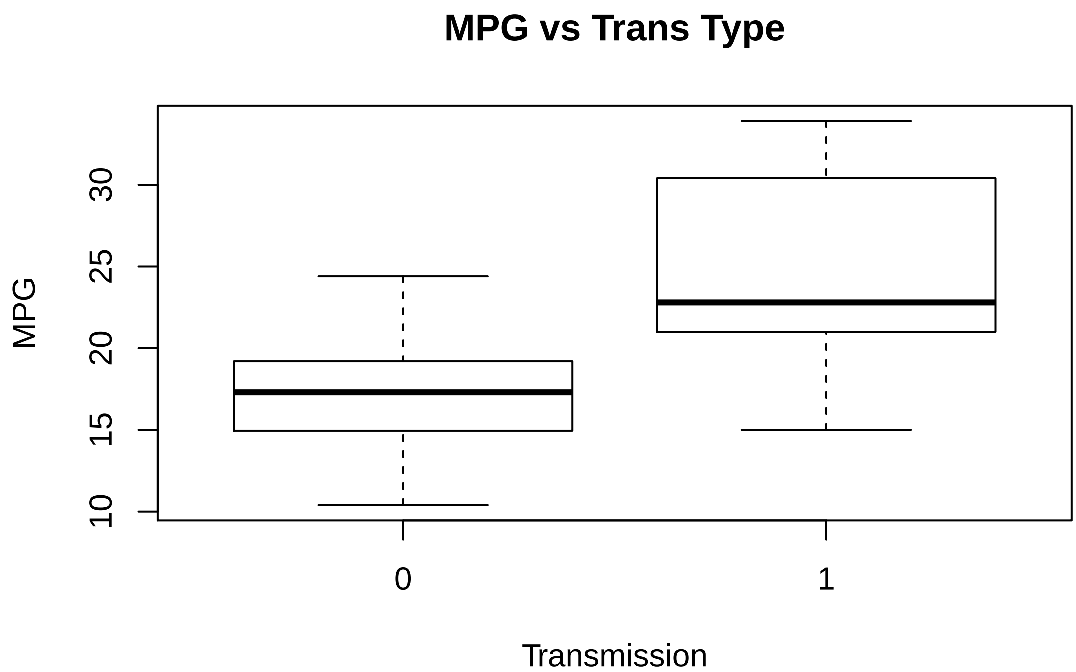
```
##
## Call:
## glm(formula = mpg ~ as.factor(cyl) + as.factor(am) + hp + wt -
##     1, data = mtcars)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9387  -1.2560  -0.4013   1.1253   5.0513
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## as.factor(cyl)4 33.70832    2.60489  12.940 7.73e-13 ***
## as.factor(cyl)6 30.67698    3.10835   9.869 2.79e-10 ***
## as.factor(cyl)8 31.54465    3.88461   8.120 1.34e-08 ***
## as.factor(am)1   1.80921    1.39630   1.296  0.20646
## hp              -0.03211    0.01369  -2.345  0.02693 *
## wt              -2.49683    0.88559  -2.819  0.00908 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 5.808677)
##
##     Null deviance: 14042.31  on 32  degrees of freedom
## Residual deviance:   151.03  on 26  degrees of freedom
## AIC: 154.47
##
## Number of Fisher Scoring iterations: 2
```

As our model reveals, there is a difference of 1.8 mpg accounted by the transmission type. However, since the standard error is large, which does not pass the t test (t value = 0.2), the dependence cannot be interpreted as significant.

## Appendix:

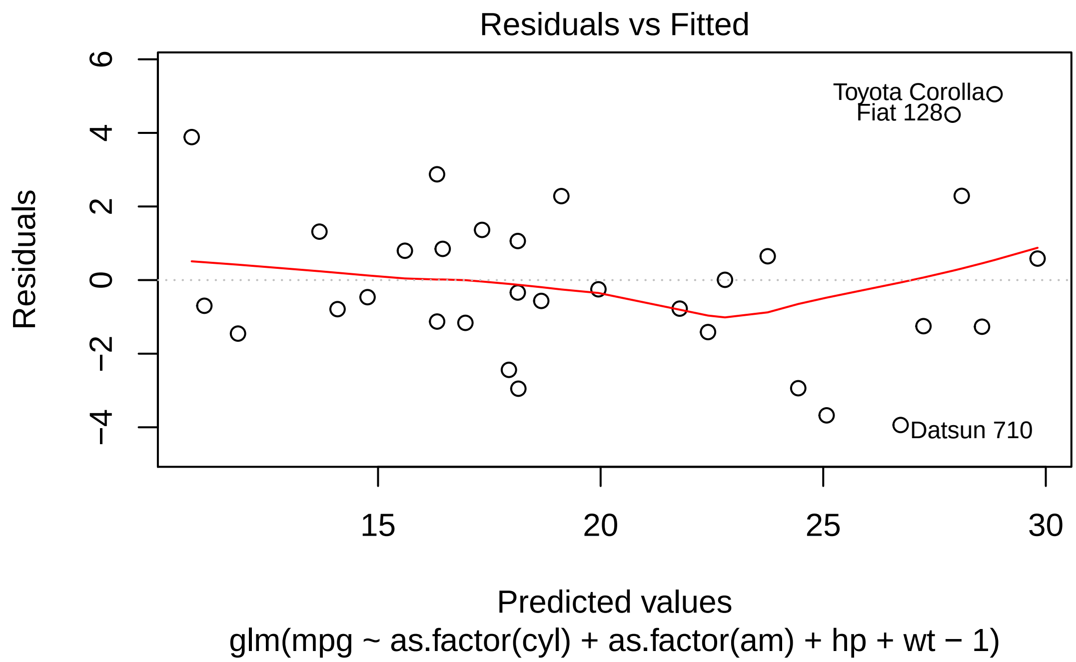
Plot of mpg statistics for auto and manual transmission cars. On the x-axis, 0 and 1 represents manual and auto transmission, respectively.

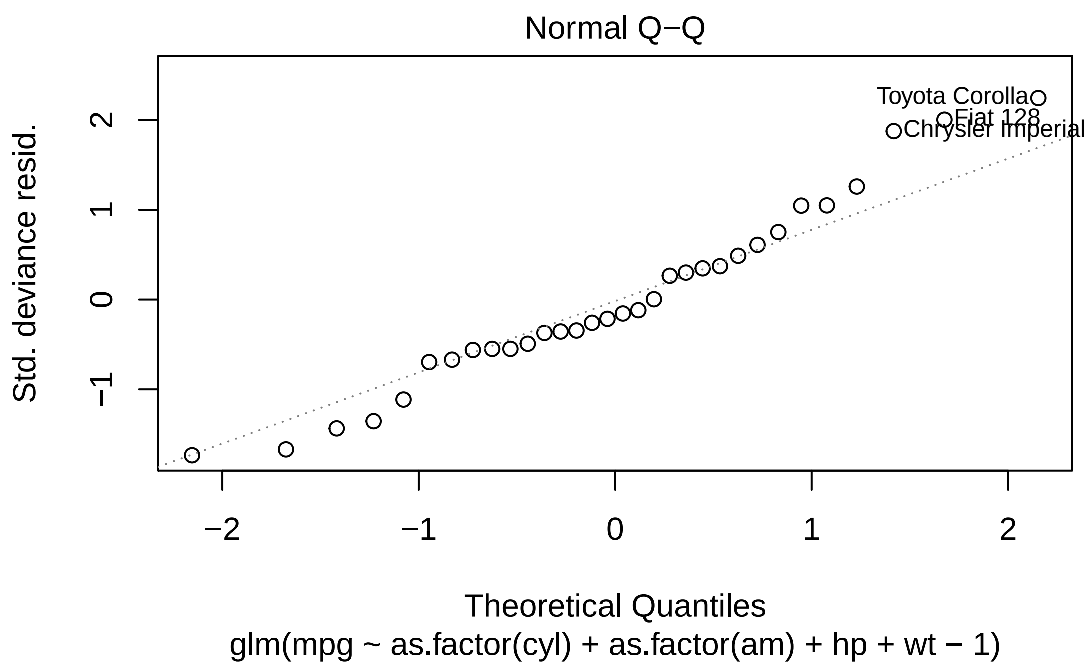
```r
boxplot(mpg ~ am, data = mtcars, xlab = "Transmission", ylab = "MPG", main="MPG vs Trans Type")
```

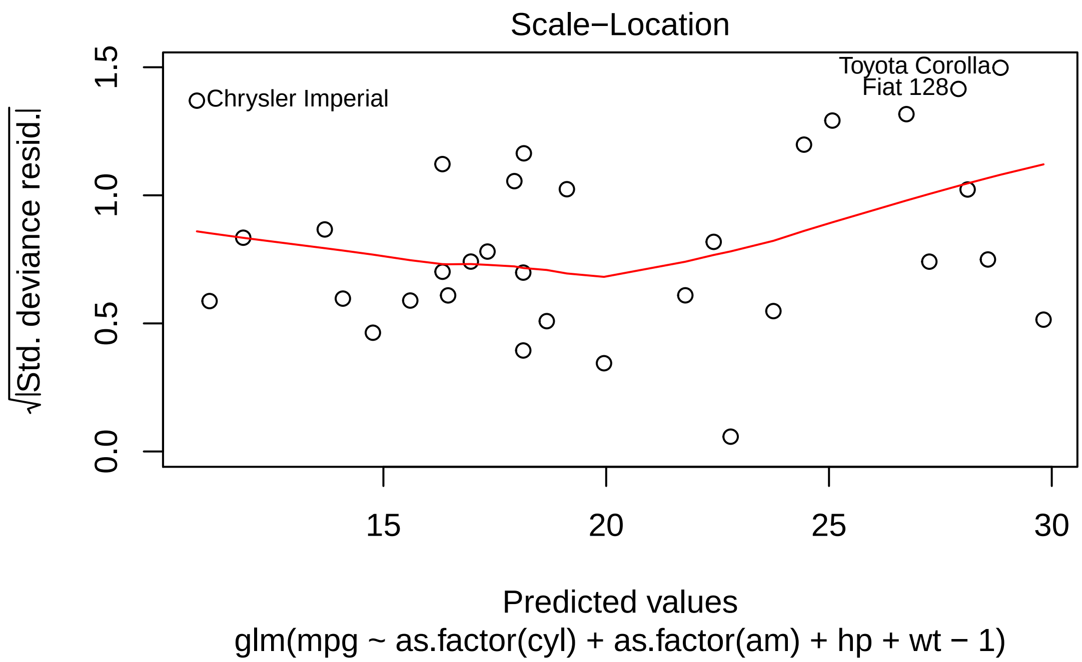


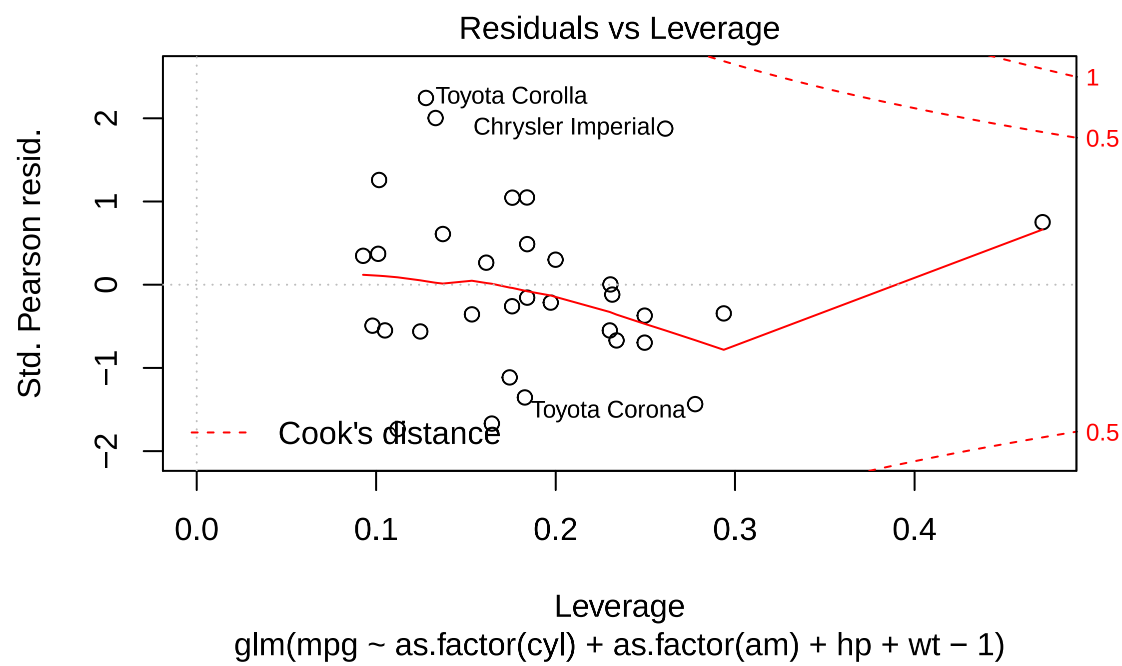

Plot of selected model fitting:

```
plot(fit2)
```


Residuals vs Fitted
Residuals
Predicted values
glm(mpg ~ as.factor(cyl) + as.factor(am) + hp + wt − 1)
Toyota Corolla
Fiat 128
Datsun 710
6
4
2
0
−2
−4
15
20
25
30


Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
glm(mpg ~ as.factor(cyl) + as.factor(am) + hp + wt − 1)
Toyota Corolla
Fiat 128
Chrysler Imperial
2
1
0
−1
−2
−1
0
1
2


Scale−Location
√|Std. deviance resid.|
0.0
0.5
1.0
1.5
15
20
25
30
Chrysler Imperial
Toyota Corolla
Fiat 128
Predicted values
glm(mpg ~ as.factor(cyl) + as.factor(am) + hp + wt − 1)


Residuals vs Leverage
Std. Pearson resid.
−2
−1
0
1
2
0.0
0.1
0.2
0.3
0.4
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5
Leverage
glm(mpg ~ as.factor(cyl) + as.factor(am) + hp + wt − 1)