Influence of transmission type on MPG

Executive summary

There is a statistically significant difference between mpg values for different types of transmission. Analysis based on linear models shows that automatic transmission gives 2.9358 more miles per gallon comparing to manual transmission.

Exploratory analysis

Loading data

```
library(datasets)
data(mtcars)

mtcars$am<-factor(mtcars$am)
```

We see on graph (Appendix 1) considerable difference among mpg means for two types of transmission. Let's see if this difference is statisitically significant.

```
fit<-lm(mpg~am, mtcars)
summary(fit)
```

```
##
## 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 ***
## am1            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
```

Intercept here represents manual transmission, while automatic gives significant (at p-value<0.05) increase. But this model explains only 36% of variance, besides chances are the coefficient we get is not true because of the hidden correlated variables. So we gonna add some predictors to the model.

Regression analysis

Let's begin adding all the variables as regressors.

```
fit<-lm(mpg~., mtcars)
summary(fit)$coef
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am1          2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

Now we'll be removing predictors one by one according to the highest p-values until we get meaningful model.

```r
fit<-lm(mpg~.-cyl, mtcars)
fit<-lm(mpg~.-cyl-vs, mtcars)
fit<-lm(mpg~.-cyl-vs-gear, mtcars)
fit<-lm(mpg~.-cyl-vs-gear-carb, mtcars)
fit<-lm(mpg~.-cyl-vs-gear-carb-drat, mtcars)
fit<-lm(mpg~.-cyl-vs-gear-carb-drat-disp, mtcars)
fit<-lm(mpg~.-cyl-vs-gear-carb-drat-disp-hp, mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ . - cyl - vs - gear - carb - drat - disp -
##     hp, 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 ***
## am1           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
```

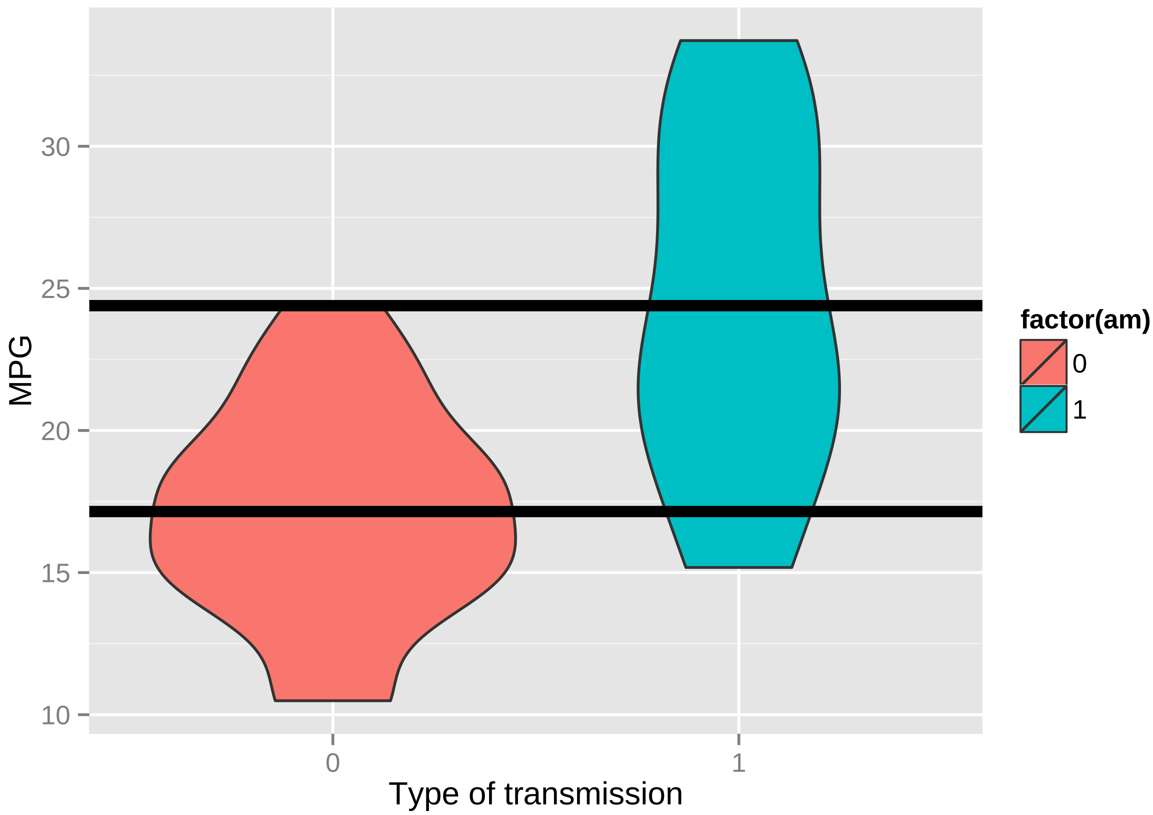
Now we have model with wt, qsec and am as regressors, wich explains about 85% of variance. We see that automatic transmission gives now 2.9358 increase in mpg with other predictors keeping fixed.

Residuals and outliers diagnostics

On the residuals plot (Appendix 2) we see no pattern, residuals are approximately identically distributed above and below 0, which means our model is quite good. On the leverage vs residuals plot (Appendix 3) we see no signifaicant leverage for any point in our data, which means outliers don't have much influence on our model.

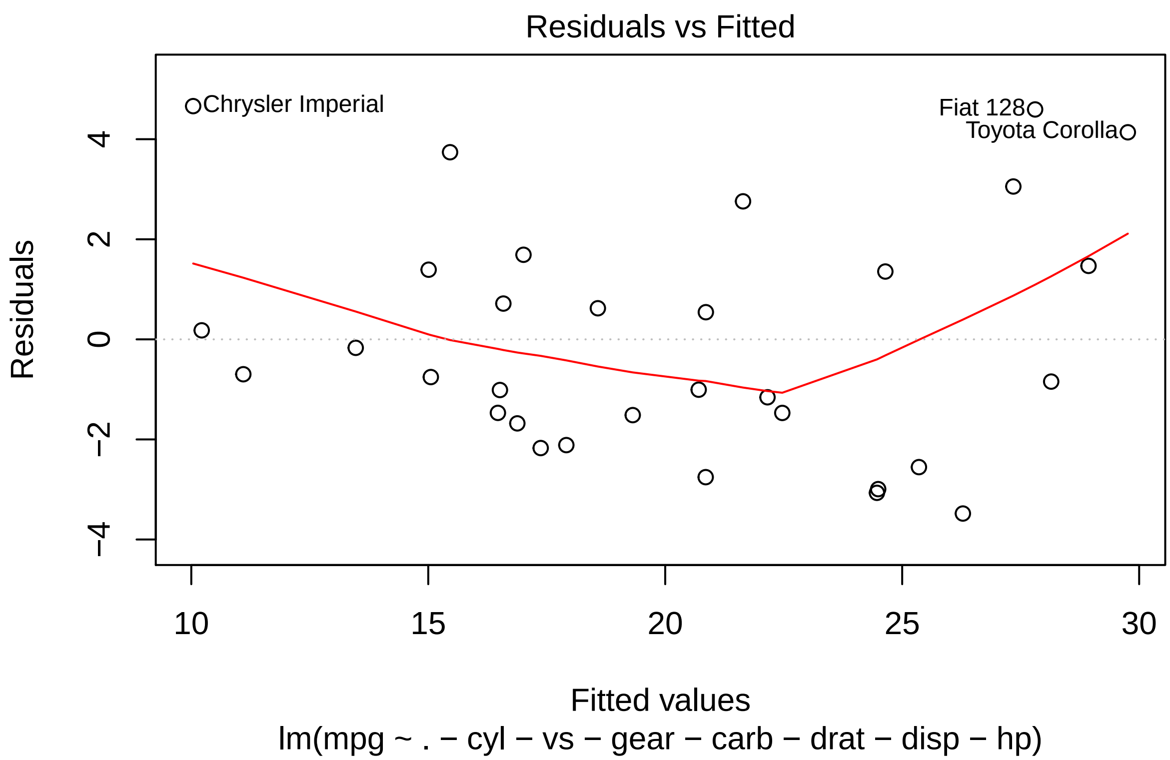
Appendix 1

```
library(ggplot2)
g<-ggplot(mtcars, aes(x=factor(am), y=mpg, fill=factor(am)))+
geom_violin()+
xlab("Type of transmission")+ylab("MPG")+
geom_abline(intercept=mean(mtcars$mpg[mtcars$am==0]), slope=0, size=2)+
geom_abline(intercept=mean(mtcars$mpg[mtcars$am==1]), slope=0, size=2)
g
```



Appendix 2

```
plot(fit,which=1)
```



Appendix 3

```
plot(fit,which=5)
```