Regression Models Course Project

Vincent MARIN

28 juillet 2016

Executive Summary

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

Based on a panel of thirty-two 1973-74 cars, manual transmission is better for MPG than automatic by 2,94 miles per gallon. You should consider first the weight of the vehicule, an increase of 1000lb will impact negatively the gas mileage by 3,92 miles per gallon, and then take into account the 1/4 mile time (+1,23).

Exploratory Data Analysis

```
data(mtcars)
head(mtcars)
```

This dataset has 32 observations on 11 variables. vs and am should be moddeled as categorical variables.

```
mtcars$am <- as.factor(mtcars$am)
mtcars$vs <- as.factor(mtcars$vs)
```

To simplify the analysis, two dataframes are created : the first one with only automatic transmission associated variables values and the second one with manual transmission associated variables alues.

```
mtcars0 <- mtcars[which(mtcars$am=='0'),]
mtcars1 <- mtcars[which(mtcars$am=='1'),]
```

Let's take a look at mpg values.

```
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.42   19.20   20.09   22.80   33.90
```

```
summary(mtcars0$mpg)
```

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

```
summary(mtcars1$mpg)
```

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

Automatic cars have a worst mpg ratio than manual ones.

Regression models

The MPG ratio is fitted to the transmission type

```
fit0 <- lm(mpg ~ am, data=mtcars)
```

The MPG ratio is fitted against all the other variables

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

In order to get the best model, we use the step function. This algorithm will select the model's variables which will provide the lowest AIC by dropping variables one at a time.

```
fit2 <- step((fit1),direction="backward",trace=0)
```

```
c(summary(fit0)$adj.r.squared,summary(fit1)$adj.r.squared,summary(fit2)$adj.r.squared)
```

```
## [1] 0.3384589 0.8066423 0.8335561
```

The first, the second and the last models explain respectively 34%, 81% and 84% of the variance of the MPG variable

```
c(summary(fit0)$sigma,summary(fit1)$sigma,summary(fit2)$sigma)
```

```
## [1] 4.902029 2.650197 2.458846
```

The next model has always a smaller residual standard error.

```
summary(fit0)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

am1 is significant at a 0,05 level. On average a car has a 17,15 gas mileage. A car with a manual transmission has a better ratio (17,15 + 7,25 = 24,30).

```
summary(fit1)$coefficients
```

```
##                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
## vs1          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
```

None of the coefficients are significant at a 0,05 level. If all the other variables remain constant, a car with a manual transmission has a better gas mileage (12,30 + 2,52 = 14,82).

```
summary(fit2)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am1          2.935837  1.4109045  2.080819 4.671551e-02
```

Only the intercept is not significant at at a 0,05 level. Manual cars still have a better gas mileage. An increase of 1000lb has a bigger impact (-3,91) on the MPG than the transmission type (2,94).

Residuals

```
bptest(fit2)$p.value
```

```
##        BP
## 0.1028523
```

The points of the first plot follow the red line so we can tell that the moddel is correctly fitted. The QQ plot points fall on the line : residuals are normally distributed. Residuals are equally spread along the ranges of predictors. We cannot reject the null hypothesis of homoscedasticity based on the Breusch-Pagan even if the red line is not horizontal. Outliers such as Fiat 128 or Merc 230 do not fall under the Cook's distance and then are not considered influencial.

Inference

```
t.test(mpg ~ am, data=mtcars)$p.value
```

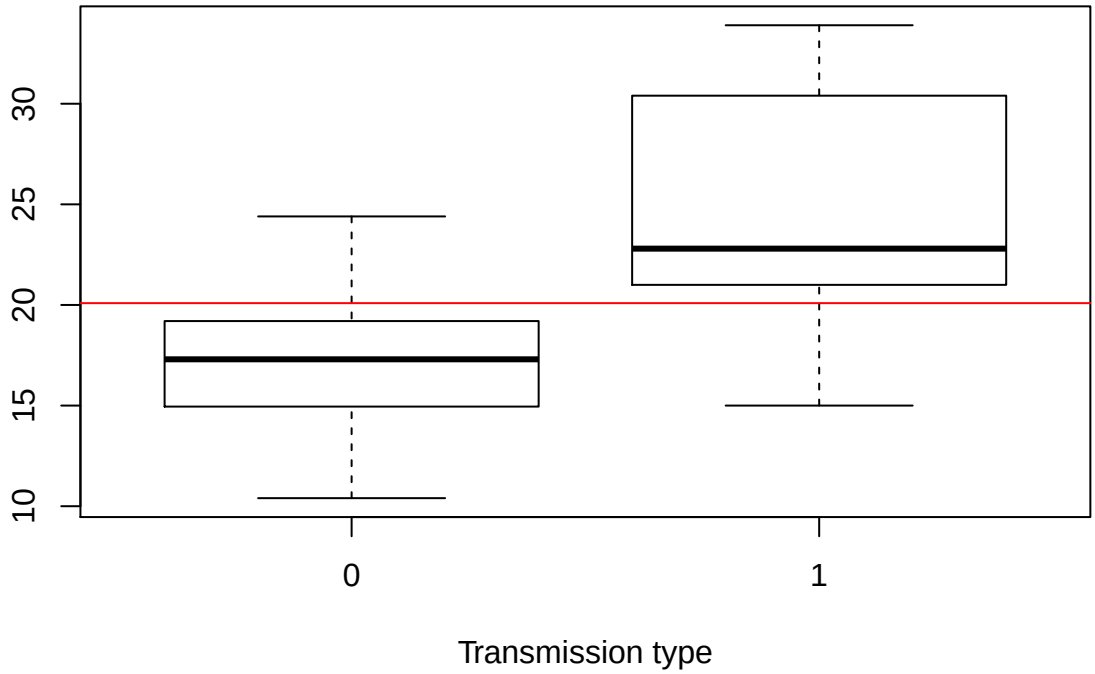
```
## [1] 0.001373638
```

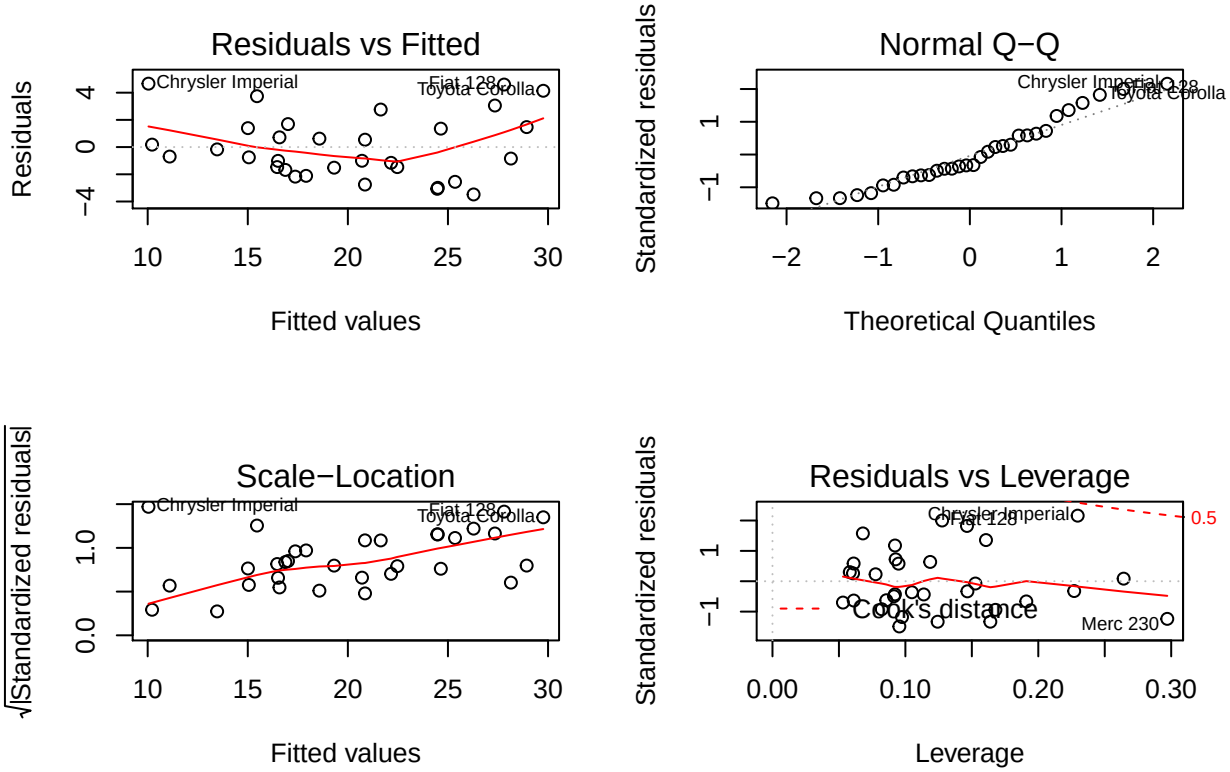
```
t.test(mpg ~ am, data=mtcars)$estimate
```

```
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

With a p-value of 0,0013 we reject the null hypothesis. Automatic and manual cars are significantly different from each other.

Appendix




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