# Automatic or Mechanical, which one to choose?

*Tomas Fuenzalida*

*4 de mayo de 2020*

## Introduction

In this report, a complete analysis of the "mtcars" data that is part of the R program will be performed. In this report, we will fit a linear regression model to explain the difference between the mpg (miles per gallon) of mechanical and manual automobiles.

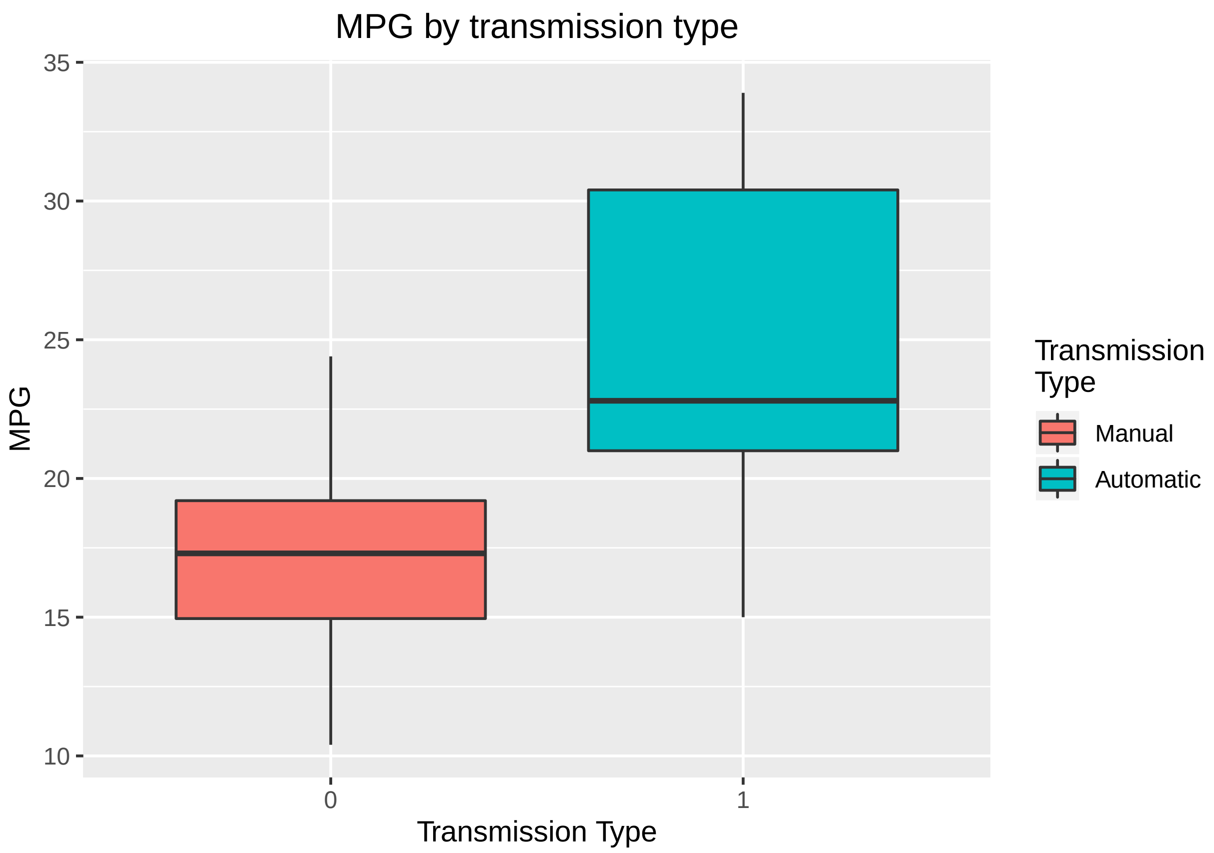
## Summary

Performing a full scan to find the difference between mechanical and manual cars in mpg, it was found that the relevant variables for this study were weight (wt), transmission type (am), and car acceleration (qsec). With this adjusted model it was found that for automatic models we reach 3mpg on average than for mechanical models

First, an exploratory analysis of the data is carried out to understand these in order to obtain a good starting point

```
##      vars  n   mean     sd median trimmed    mad   min    max  range  skew
## mpg     1 32  20.09   6.03  19.20   19.70   5.41 10.40  33.90  23.50  0.61
## cyl*    2 32   2.09   0.89   2.00    2.12   1.48  1.00   3.00   2.00 -0.17
## disp    3 32 230.72 123.94 196.30  222.52 140.48 71.10 472.00 400.90  0.38
## hp      4 32 146.69  68.56 123.00  141.19  77.10 52.00 335.00 283.00  0.73
## drat    5 32   3.60   0.53   3.70    3.58   0.70  2.76   4.93   2.17  0.27
## wt      6 32   3.22   0.98   3.33    3.15   0.77  1.51   5.42   3.91  0.42
## qsec    7 32  17.85   1.79  17.71   17.83   1.42 14.50  22.90   8.40  0.37
## vs*     8 32   1.44   0.50   1.00    1.42   0.00  1.00   2.00   1.00  0.24
## am*     9 32   1.41   0.50   1.00    1.38   0.00  1.00   2.00   1.00  0.36
## gear   10 32   3.69   0.74   4.00    3.62   1.48  3.00   5.00   2.00  0.53
## carb   11 32   2.81   1.62   2.00    2.65   1.48  1.00   8.00   7.00  1.05
##      kurtosis    se
## mpg     -0.37  1.07
## cyl*    -1.76  0.16
## disp    -1.21 21.91
## hp      -0.14 12.12
## drat    -0.71  0.09
## wt      -0.02  0.17
## qsec     0.34  0.32
## vs*     -2.00  0.09
## am*     -1.92  0.09
## gear    -1.07  0.13
## carb     1.26  0.29
```

It is important to see that in the data there are 3 variables that are transformed into a factor, which are included in its name *.



As can be seen in the graph above, there is a relationship between the mpg of a car and its transmission, but the possible presence of other variables that may influence the mpg of a car should not be ruled out. That is why we will start by performing a simple linear regression with all the variables to search for possible new relationships.

```
##             mpg       disp         hp        drat         wt        qsec
## mpg   1.0000000 -0.8475514 -0.7761684  0.68117191 -0.8676594  0.41868403
## disp -0.8475514  1.0000000  0.7909486 -0.71021393  0.8879799 -0.43369788
## hp   -0.7761684  0.7909486  1.0000000 -0.44875912  0.6587479 -0.70822339
## drat  0.6811719 -0.7102139 -0.4487591  1.00000000 -0.7124406  0.09120476
## wt   -0.8676594  0.8879799  0.6587479 -0.71244065  1.0000000 -0.17471588
## qsec  0.4186840 -0.4336979 -0.7082234  0.09120476 -0.1747159  1.00000000
## gear  0.4802848 -0.5555692 -0.1257043  0.69961013 -0.5832870 -0.21268223
## carb -0.5509251  0.3949769  0.7498125 -0.09078980  0.4276059 -0.65624923
##            gear       carb
## mpg   0.4802848 -0.5509251
## disp -0.5555692  0.3949769
## hp   -0.1257043  0.7498125
## drat  0.6996101 -0.0907898
## wt   -0.5832870  0.4276059
## qsec -0.2126822 -0.6562492
## gear  1.0000000  0.2740728
## carb  0.2740728  1.0000000
```

In the correlation matrix it can be seen that the variables “disp”, “hp”, “drat” and “wt” have high correlation with each other, so it is important to take this into account when fitting a linear regression model

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

```
##           GVIF Df GVIF^(1/(2*Df))
## cyl  36.345193  2        2.455341
## disp 21.631435  1        4.650961
## hp   16.078598  1        4.009813
## drat  3.736566  1        1.933020
## wt   15.182209  1        3.896435
## qsec  7.739648  1        2.782022
## vs    6.101654  1        2.470153
## am    4.653616  1        2.157224
## gear  5.371501  1        2.317650
## carb 10.775733  1        3.282641
```

Taking into account the VIF and the characteristics of the variables, we started our linear regression model including the “am”, “hp” and “qsec” vairbals, hoping that they could fit better.

```
fit2 <- lm(mpg ~ am + hp + qsec, data = mtcars)
fit3 <- update(fit2, mpg ~ am + hp + qsec + cyl)
fit4 <- update(fit3, mpg ~ am + hp + qsec + cyl + wt)
anova(fit2,fit3,fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + hp + qsec
## Model 2: mpg ~ am + hp + qsec + cyl
## Model 3: mpg ~ am + hp + qsec + cyl + wt
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     28 238.56
## 2     26 196.91  2    41.647 3.6156 0.041762 *
## 3     25 143.98  1    52.932 9.1907 0.005596 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit4)$coeff
```

```
##                Estimate  Std. Error     t value    Pr(>|t|)
## (Intercept) 21.57616508 11.27271090  1.91401742 0.067132106
## am1          2.83269554  1.67019936  1.69602241 0.102301456
## hp          -0.02480654  0.01514918 -1.63748374 0.114056916
## qsec         0.61917236  0.55987340  1.10591494 0.279292531
## cyl6        -1.90949640  1.72992011 -1.10380611 0.280188887
## cyl8        -0.22716240  2.87047074 -0.07913768 0.937553171
## wt          -2.96274129  0.97728243 -3.03161217 0.005595827
```

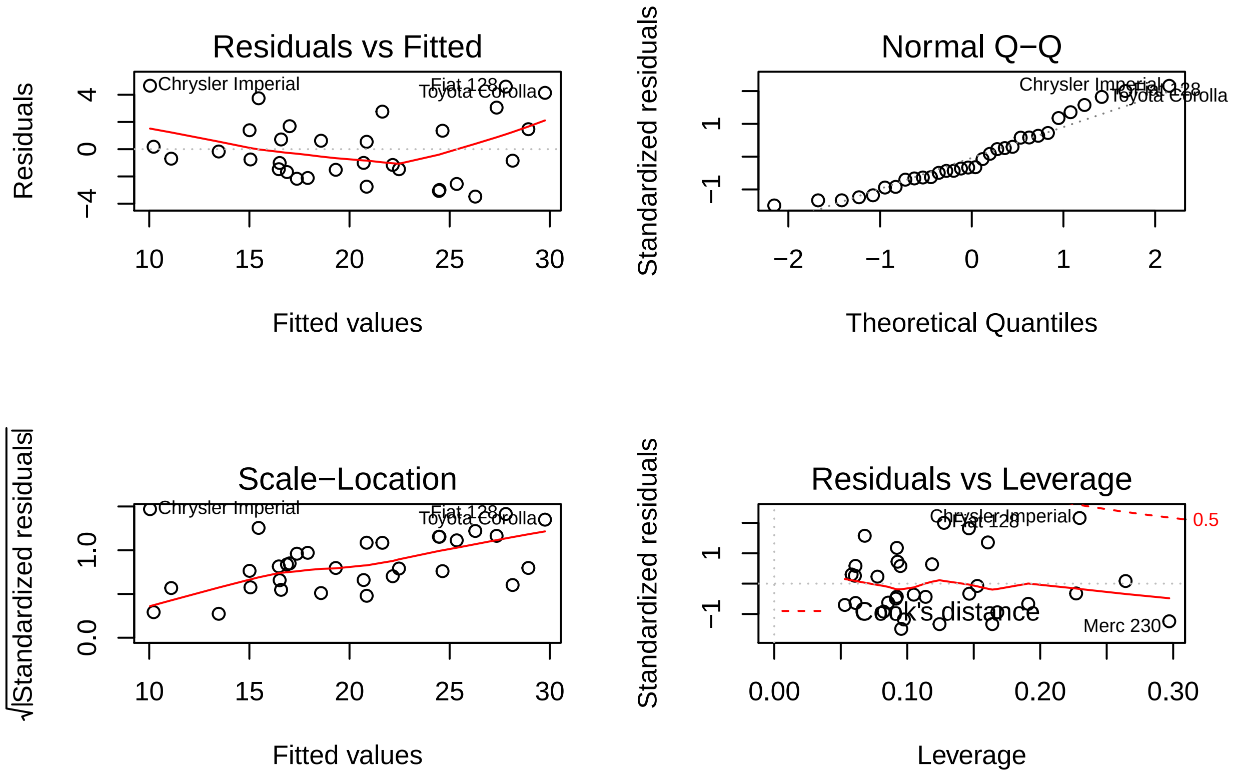
Although the ANOVA function suggests that we include the variable “cyl”, the coefficients when performing the regression give us that “cyl” is not very significant, so we will remove it from the regression.

```
fit6 <- lm(mpg ~ am + qsec + wt, data = mtcars)
summary(fit6)$coeff
```

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

Finally iterating we come to the conclusion that mpg is described by “am”, “qsec” and “wt”.

```
par(mfrow = c(2,2))
plot(fit6)
```



In the first plot we can see that there is no clear pattern so it fits well The second plot suggests that the errors conform to normal apparently. The third we find points apparently randomly distributed, which suggests a homocedastic and bias-free model Finally, the fourth is that there are no points that have such a significant impact on the regression, so it confers a good fit.

The model gives us that there is a difference of approximately 3 mpg of difference between automatic and manual models. Also that the more acceleration the more MPG at a ratio of 1.22. Finally every 1000lbs more about 4 MPG is lost