Fuel Comsumption Analisys

Joaquin Solino Perez

June, 2021

Summary

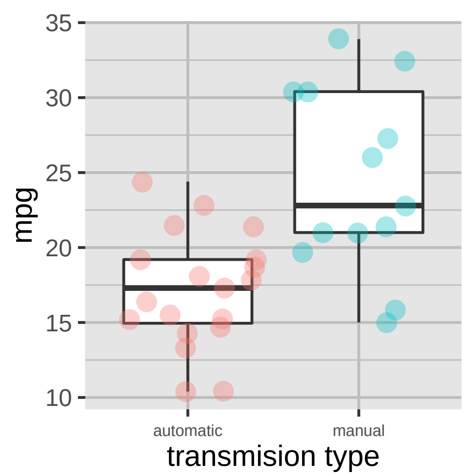
In this document we will try to anwer the question of wich transmission type is better for fuel consumption in cars. It may seen like an easy question to answer, but it is not so simple. There are many variables that affect fuel consumption. Transmision type is a relevant variable, but not the only one. In this document we'll try to explain wich are the most influential variables and the numerical impact in fuel consumption in cars.

Car Road Tests

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). You can see the full description and a summary of the data in **Apendix A**.

Wich type of transmission is better to fuel consumption?

The variable related to fuel consumption is (`mpg`). Let's 'see a boxplot with the `mpg` vs transmission type (`am`):



You can conclude that see transmision cars performs, in general, greater mpg. To check tha the difference is statistically significant, we make a `t.test`

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group automatic    mean in group manual
##                17.14737                24.39231
```

Looking for transmission type uniquelly, manual transmision has a better fuel consumptuion vs automatic, and the difference is significant (*p-value* $= 0.00137 < 0.05$) and with a mean difference of 7.24 miles per gallon.

In fact, the previous analysis is not complete. This is because there are many variables that affect on the `mpg` (see in Apendix A. If we fit a linear model to estimate mpg only from transmission type, you find that is explains a low variance.

```
sum_fit_am <- summary(lm(mpg ~ am, data = my_cars))
sum_fit_am$coefficients
```

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

```
sum_fit_am$adj.r.squared
```

```
## [1] 0.3384589
```

> 0.338 is a poor value of the We have to conclude than transmision type separatelly (`am` variable) is not enough to measure the difference in mpg

Quantify impact of transmission type on fuel consumption

There are more variables that are relevant. In *Apendix A - Model Selection* we try different models to select the best balance (that is: maximize adjusted R^2 with the least estimated error in the model coefficients and normality of residuals). The result is the model that includes transmission type (am), the weight (wt) and the time to run 1 1/4 miles (qsec).

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec, data = my_cars)
##
## 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
## ammanual      2.9358     1.4109   2.081 0.046716 *
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ---
## 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
```

This model, can explain over 83% of variance of mpg. You can see that am has a positive coefficient, so manual tranmision increases mpg in 2.9358372 that is significantly lower than the first calculus with only transmission type.

Apendix A

Structure of `mtcars` data frame

`mtcars` is a data frame with 32 observations on 11 variables. Next you hava the description

#	Var	Type	Description
1	`mpg`	Numeric	Miles/(US) gallon
2	`cyl`	Numeric	Number of cylinders
3	`disp`	Numeric	Displacement (cu.in.)
4	`hp`	Numeric	Gross horsepower
5	`drat`	Numeric	Rear axle ratio
6	`wt`	Numeric	Weight (1000 lbs)
7	`qsec`	Numeric	1 1/4 mile time
8	`vs`	Factor	Engine (0 = V-shaped, 1 = straight)
9	`am`	Factor	Transmission (0 = automatic, 1 = manual)
10	`gear`	Numeric	Number of forward gears
11	`carb`	Numeric	Number of carburetors

Summary of data

Next you have a summary of the dataset:

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec        vs             am
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   V:18   manual   :19
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   S:14   automatic:13
##  Median :3.695   Median :3.325   Median :17.71
##  Mean   :3.597   Mean   :3.217   Mean   :17.85
```

```
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90
##  Max.   :4.930   Max.   :5.424   Max.   :22.90
##       gear            carb
##  Min.   :3.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:2.000
##  Median :4.000   Median :2.000
##  Mean   :3.688   Mean   :2.812
##  3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :8.000
```

For each numeric variable, the summary shows the minimum, maximum, mean, 1st Quantile (lowest 25%), median (sorted 50%) and 3rd Quantile (top 25%). For each Factor variable you'll find the numer of observations of each level.

Pair analysis

Next you can see a matrix graphic that shows pair relation between variables.

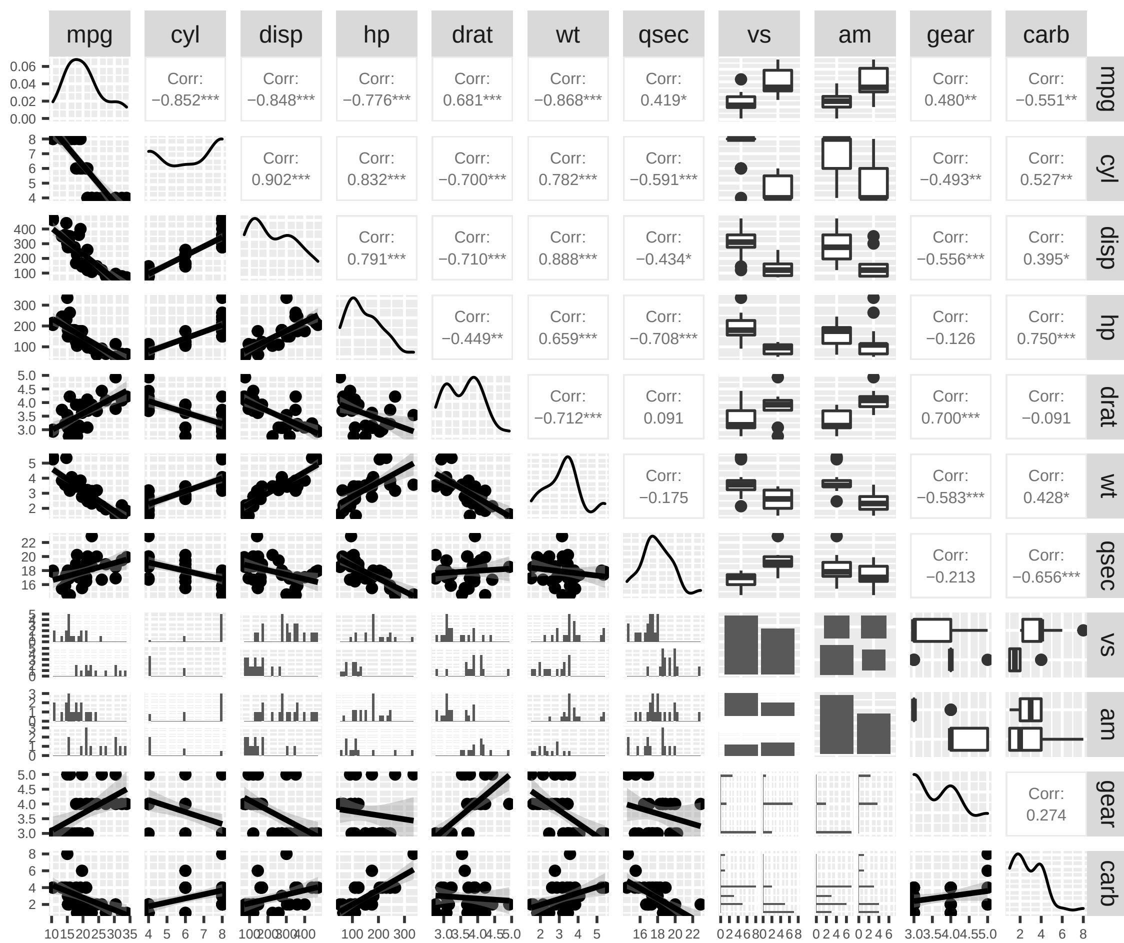


Figure 1: Pair relations - mtcars

Model selection

```
##                                                             adj_r_squared
## am + wt + qsec + disp + hp                                      0.8375334
## am + wt + qsec                                                  0.8335561
## wt + qsec + disp + hp                                           0.8107212
## cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb      0.8066423
## am + qsec                                                       0.6652425
## am                                                              0.3384589
```

With `am` we only explain 33% of variety. Including other predictors, we can achive over 83%, wich es much better. Now we check the estandar error of Coefficients for the TOP-3 R^2 models:

```
## am + wt + qsec + disp + hp

## (Intercept) amautomatic          wt        qsec        disp          hp
##  9.74079485  1.48578009  1.19409972  0.47543287  0.01060333  0.01450469

## am + wt + qsec

## (Intercept) amautomatic          wt        qsec
##   6.9595930   1.4109045   0.7112016   0.2886696

## wt + qsec + disp + hp

## (Intercept)          wt        qsec        disp          hp
##  8.63903219  1.26585131  0.46649316  0.01073767  0.01561305
```

The best balanced model would be the one with variables: am, wt and qsec. Lets check the residual plots

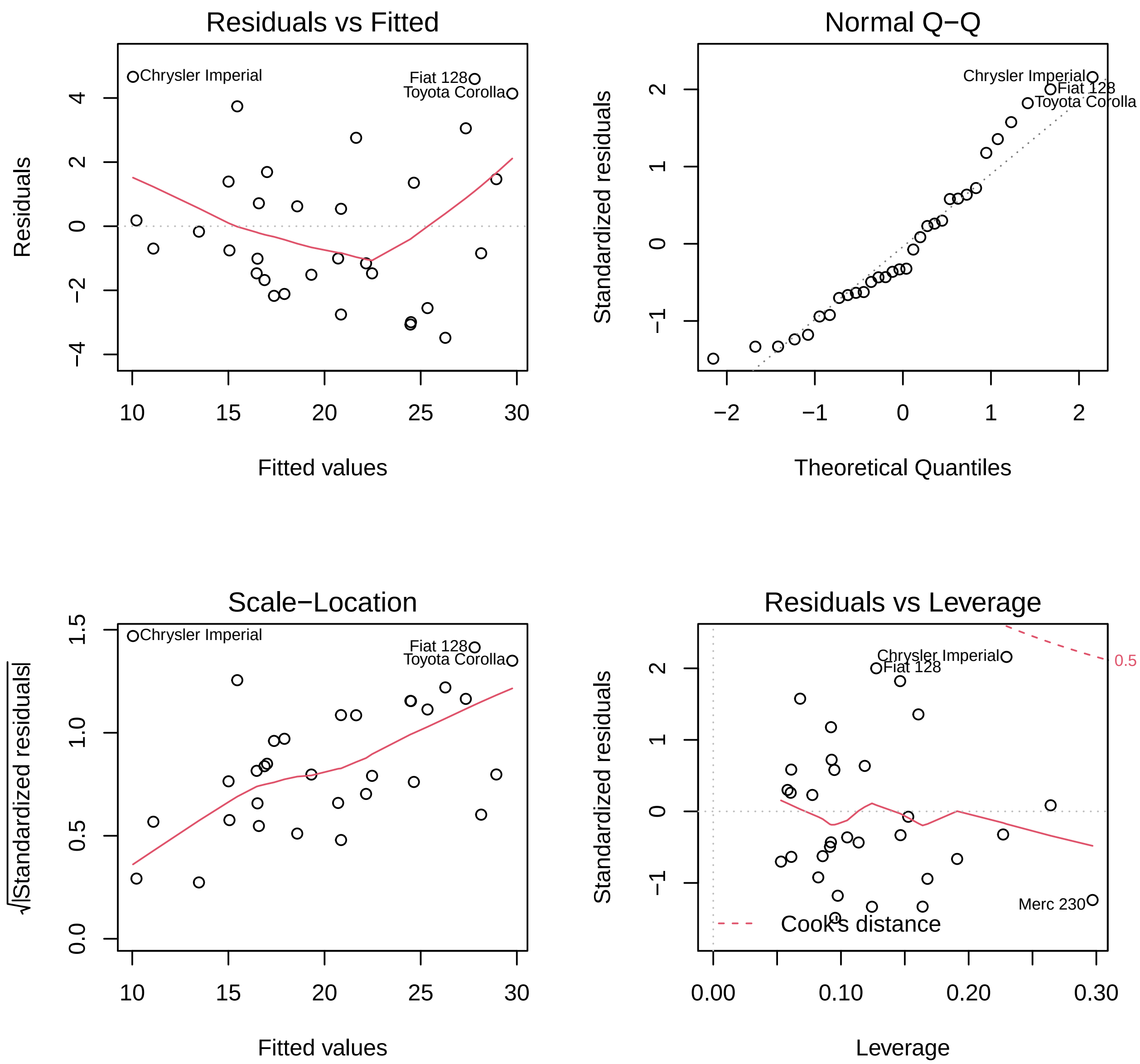


Figure 2: Model mpg vs am, wt and qsec