Study of efficiency of cars based on type of transmission

Executive summary

This document presents a study of the relationship between different variables and the efficiency of cars, represented by the variable *mpg* (miles per gallon). Specially, we are interested in which type of cars are more efficient, automatic or manual, represented by the variable *am*. In the analysis behind we will see that, on average, manual cars are more efficient in about 7.24 mpg, although this difference can be affected by other variables.

Exploratory analysis

We plot the relations between all the variables. We distinguish between automatic transmission (in green) and manual transmission (in blue). This figure is shown in the appendix as *Figure 1*.

More in detail we explore the relation between *mpg* and *am* (the type of transmission, 0-automatic/1-manual). This figure is shown in the appendix as *Figure 2*. As it can be seen, it seems to be an important effect of the type of transmission in the efficiency of a car. In the next section we are going to measure this effect.

Regression analysis

The simplest possible regression model is *mpg* vs *am*. We perform this regression model:

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

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## am             7.245      1.764   4.106 2.850e-04
```

In this case the *intercept* represents the average *mpg* when *am = 0*, in other words, the average efficiency for automatic cars. The coefficient of *am* represents the average difference of *mpg* between manual and automatic cars. Both coefficient are significatives with a significance level of 0.001. The conclusion of this is that in average, manual transmision cars are more efficient than automatic cars, in about 7.24 mpg.

It is possible to fit more complex models that include more regressors. The objective is to see how they affect to the *am* coefficient.

Firstly we try with *wt* (weight of the car):

```
fit1 <- lm(mpg ~ am + wt, mtcars)
summary(fit1)$coef
```

```
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 37.32155     3.0546 12.21799 5.843e-13
## am          -0.02362     1.5456 -0.01528 9.879e-01
## wt          -5.35281     0.7882 -6.79081 1.867e-07
```

As it can be seen, the addition of *wt* greatly affects to the coefficients. Now, the intercept is more difficult to interpret. It would be the *mpg* of an automatic car if its weight was 0. The coefficient of *am* is the difference of *mpg* between a manual and an automatic car if we keep *wt* constant. Its a very small value and it is not significative. So if we take into account the weight, the type of transmission does not affect to the efficiency of the car.

Secondly we perform the regression of *mpg* vs *am* and *hp*

```r
fit2 <- lm(mpg ~ am + hp, mtcars)
summary(fit2)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 26.58491   1.425094  18.655 1.074e-17
## am           5.27709   1.079541   4.888 3.460e-05
## hp          -0.05889   0.007857  -7.495 2.920e-08
```

In this case, all the coefficients are all statistically significants. As in the regression model above, the intercept is difficult to interpret, since it is the value of *mpg* of an automatic car that has no horsepower. The coefficient of *am* represents the average difference of *mpg* between an automatic and a manual car for a fixed value of *hp*. If we compare the coefficient of *am* of the adjusted model and the unadjusted, we can see that the effect of the type of transmission in the efficiency is smaller in the adjusted model.

The last regression model we are going to study is *mpg* vs *am* and *qsec*

```r
fit3 <- lm(mpg ~ am + qsec, mtcars)
summary(fit3)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  -18.889     6.5970  -2.863 7.711e-03
## am             8.876     1.2897   6.883 1.461e-07
## qsec           1.982     0.3601   5.503 6.271e-06
```

In this case, all the coefficients are all statistically significants. The intercept is the value of *mpg* of an automatic car whose *qsec* is 0. The coefficient of *am* represents the average difference of *mpg* between an automatic and a manual car for a fixed value of *qsec*. If we compare the coefficient of *am* of the adjusted model and the unadjusted, we can see that the effect of the type of transmission in the efficiency is bigger in the adjusted model.

If we see in *Figure 1* the plots of *qsec* vs *mpg* and *hp* vs *mpg*, most of the green points (automatic cars) are behind the red line and the blue points (manual cars) are above. This is the reason why in this regression models the coefficient of *am* is significative. Nevertheless, in the plot of *wt* vs *mpg* all points are mixed.

The model we are going to select is the simplest one, that contains only *am* as regressor. As it is the one that allows us to answer the questions we are interested in.

Now, we are going to perform some diagnosis on the residuals. In *Figure 3* of the appendix we can see that the residuals have mean 0 and the variance is approximately constant. In *Figure 4* we can see that they can be approximated by a normal distribution. We can check this with the Shapiro-Wilk test:

```r
shapiro.test(fit$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit$residuals
## W = 0.9821, p-value = 0.8573
```

As the p-value is very high we fail to reject the null hypothesis and we can asume that the residuals follow a normal distribution. So the main hypothesis of the regression models are satisfied and the model can be considered appropriate.

Finally, we want to show that the coefficients of the regression model include some kind of incertitude. Here we show the 95% confidence intervals for the coefficients. Anyway, it seems that manual cars are more efficient on average.

```r
confint(fit)
```

```
##              2.5 % 97.5 %
## (Intercept) 14.851  19.44
## am           3.642  10.85
```

Appendix

Figure 1

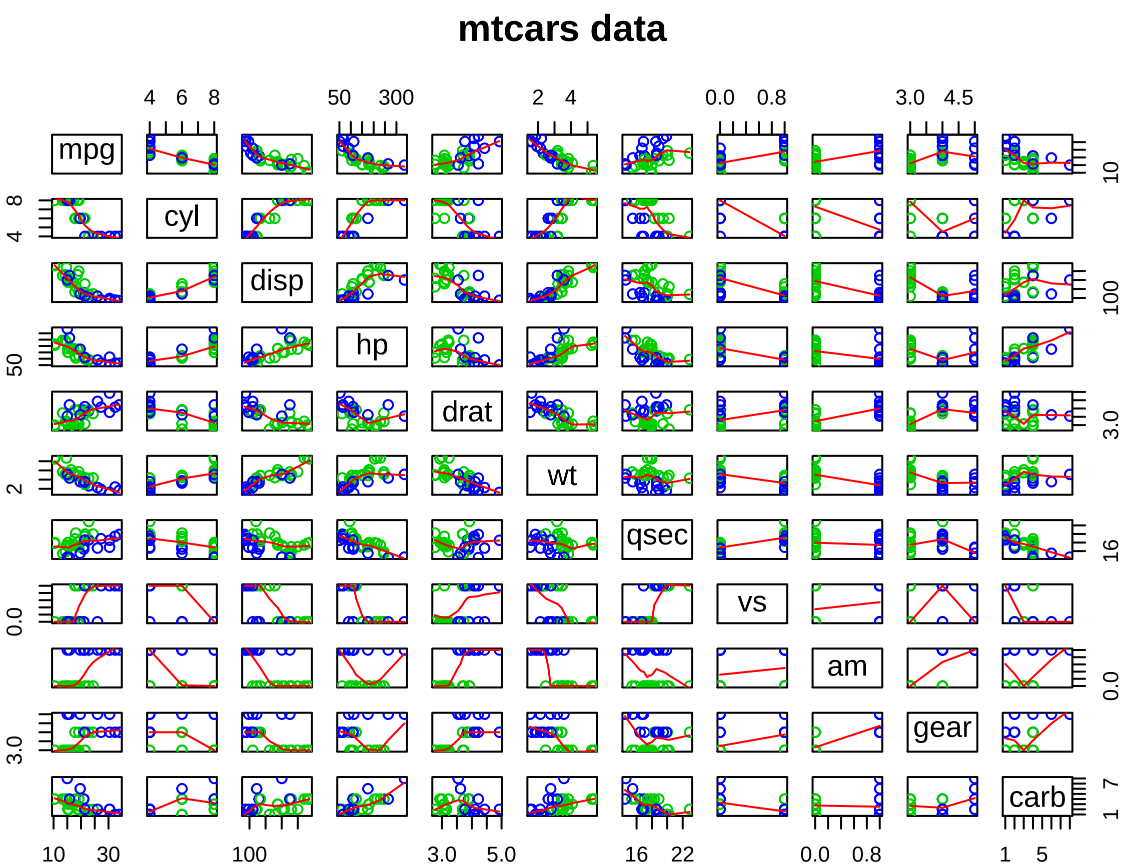


Figure 2

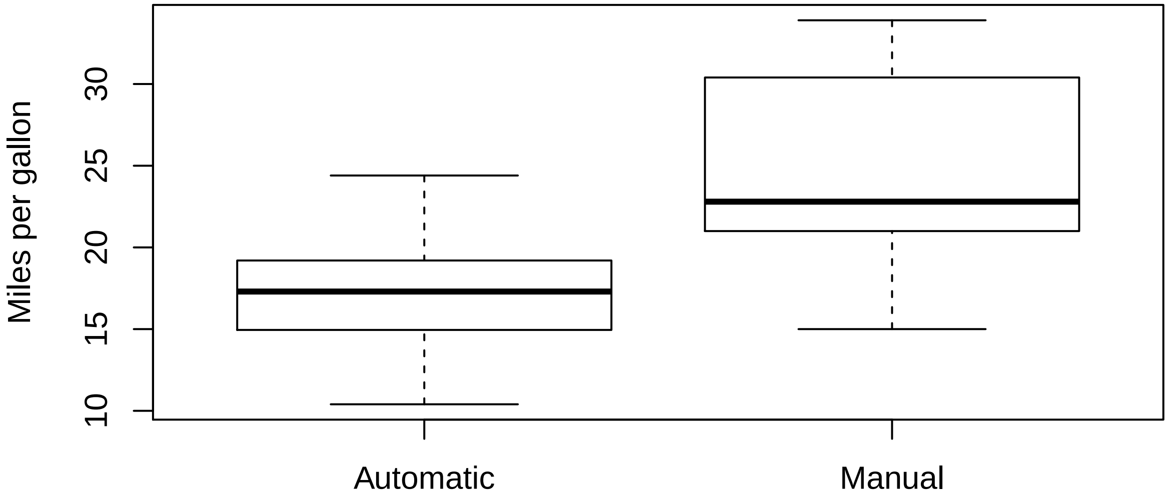


Figure 3

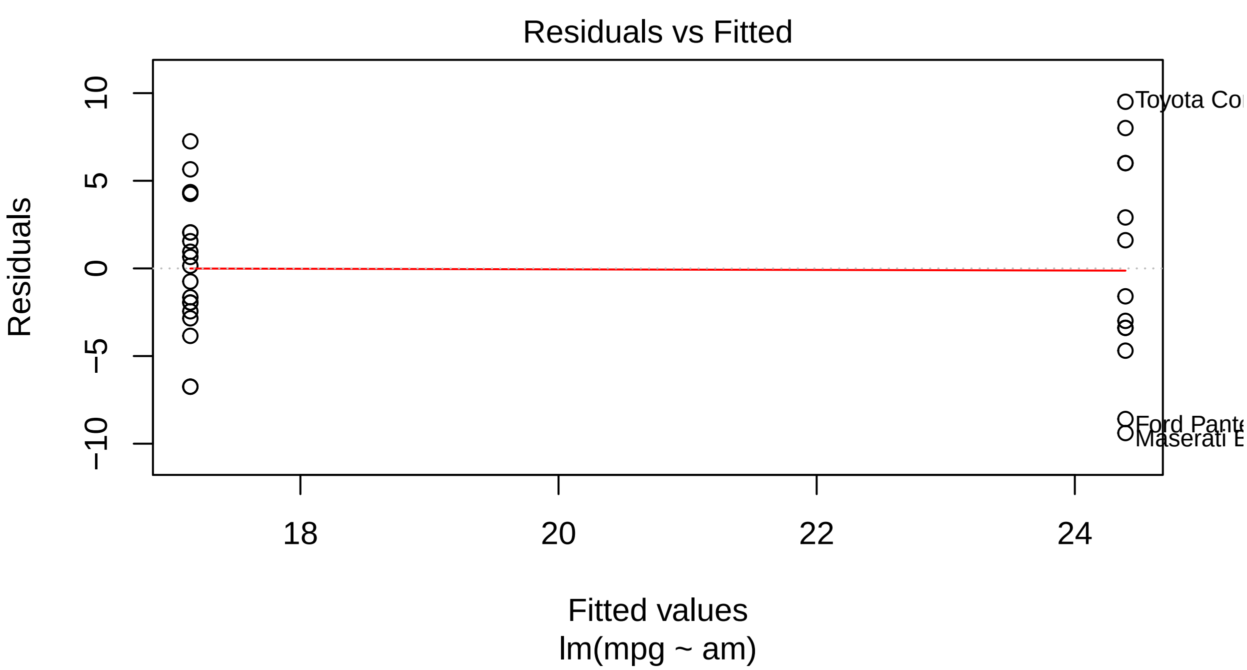
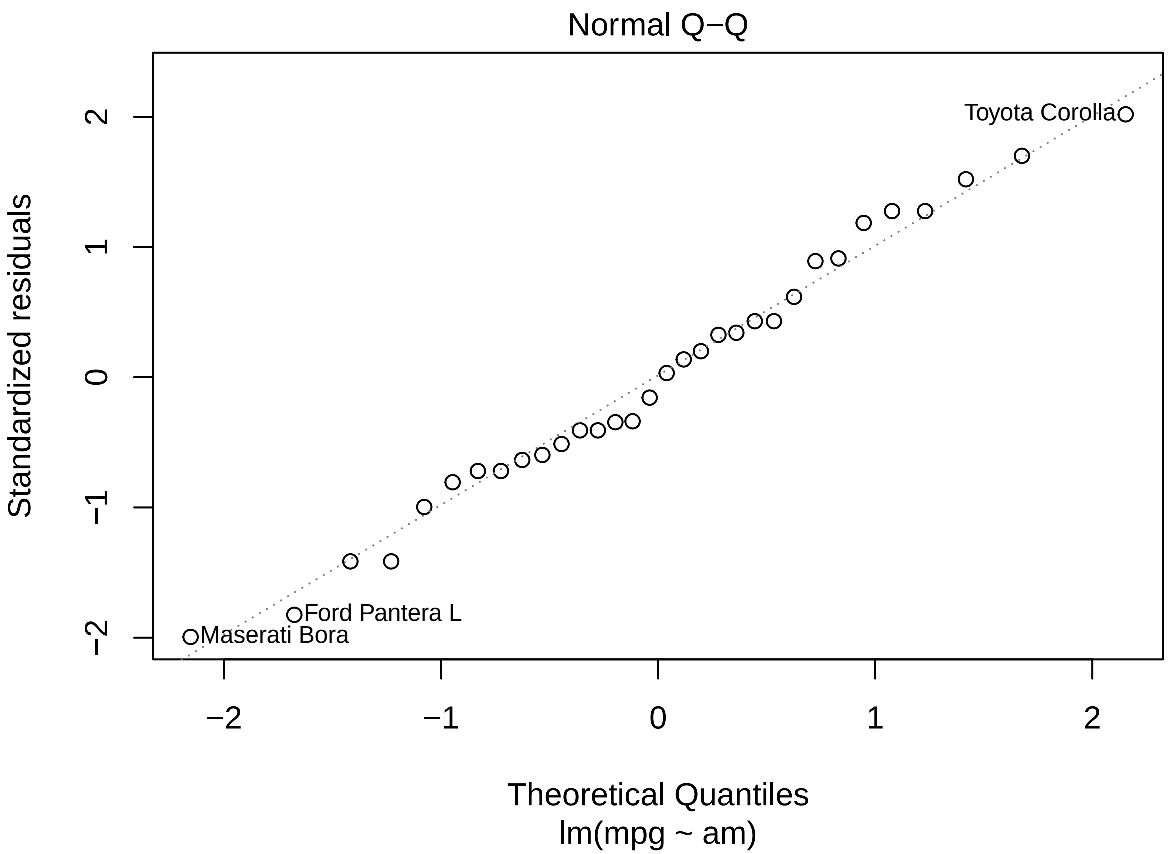


Figure 4



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ am)
Toyota Corolla
Ford Pantera L
Maserati Bora
2
1
0
−1
−2
−2
−1
0
1
2