Executive summary

In this article we study the impact of transmission type in fuel efficiency in a set of 32 cars (early 70's models). We have 13 manual transmission and 19 automatic transmission cars in the sample. A first simple analysis considering only transmission seems to indicate that manual cars are more efficient than automatic transmission cars. We have carried out a deeper analysis to determine whether transmission is relevant. It could be the case that the difference is not due to transmission but due to another underlying variable. We have used different linear models and we hace chosen the one that seems to better represent the data. We still find transmission to be a relevant factor, and that manula transmission cars are more efficient.

Is an automatic or manual transmission better for MPG?

In data set mtcars we have information for different car models (from early 70's)

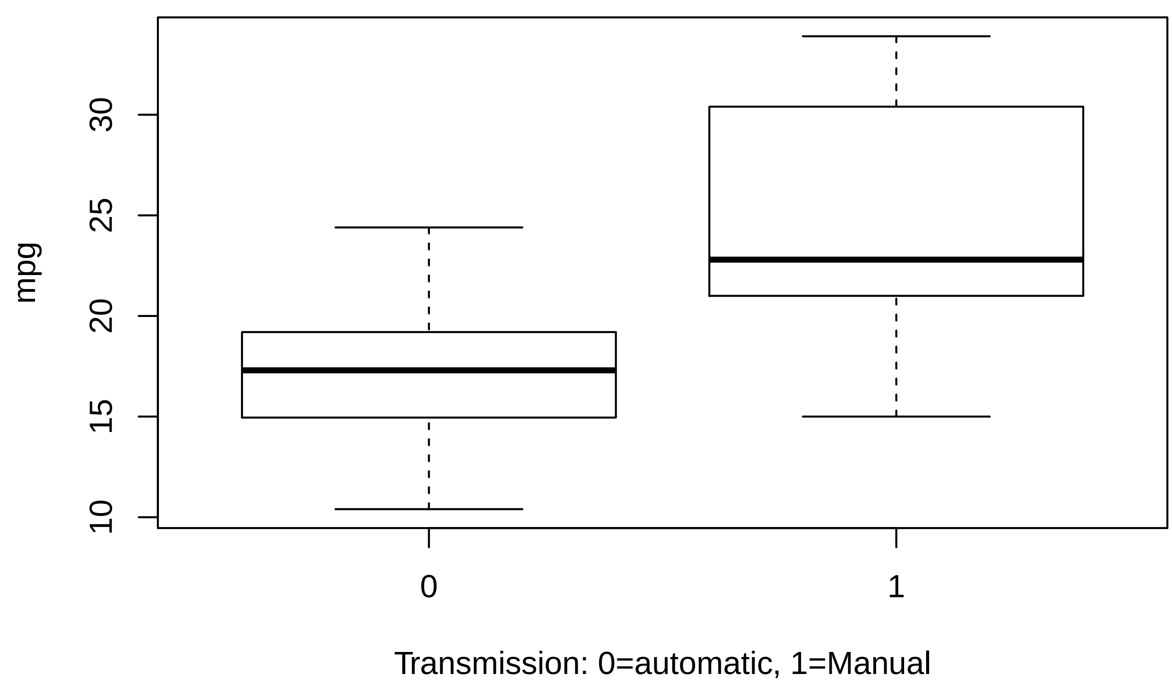
A first approach could be to compare what is the mpg for cars with automatic and manual transmission.

```
mean(mtcars$mpg[mtcars$am==1])-mean(mtcars$mpg[mtcars$am==0])
```

```
## [1] 7.244939
```

We see that manual transmission cars are in average more efficient. For a more informative analysis look at the following plot:

```
boxplot(mpg~am, data=mtcars,xlab="Transmission: 0=automatic, 1=Manual",ylab="mpg")
```



We see that cars with automatic transmission (left boxplot) tend to have lower mpg than manual transmission cars (right boxplot).

However, a deeper analysis is needed. There are other factors that have a more clear relation with mpg such as weight. It could be the case that cars with automatic transmission are heavier, being that the cause of the observed behavior.

Quantify the MPG difference between automatic and manual transmissions

As a first step, we will perform a linear regression of mpg to the rest of car characteristics.

```
lm1<-lm(mpg~.,mtcars)
summary(lm1)
```

Regression results are shown in Appendix 1. We see that the model explains 80.66% of the total mpg variance (this is the value of the adjusted R-squared). However, none of the coefficients is significant at 5% level.

We need a more parsimoniuos model, with less predictors. The next step is to identify the most relevant magnitudes. To do so we resort to information criteria. The idea is to minimize the distance of the points to our model but penalizing the presence of new predictors. The most common one is Akaike information criteria, which in R is available in the step function:

```
AIC<-step(lm1,k=log(length(mtcars$mpg)))
```

Results of this analysis can be found in Appendix 2. We see that the most relevant magnitudes in mpg are weight, transmission and 1/4 mille time. All parameters are significant at 5% level. With that model we describe 83.36% of the variance. We see that transmission is a relevant magnitude, and manual transmission cars seems to be more efficient than manual ones. From model coefficients we see that two cars with same weight and 1/4 mille time but different transmission will have different mpg, being that number higher for manual transmission. So we conclude that there is evidence that cars with manual tranmission are more efficient.

Residual analysis

We have also analyzed model residuals. See residual analysisi plots in Appendix 3.

```
lm3<-lm(mpg~wt+qsec+am,mtcars)
par(mfrow=c(2,2))
plot(lm3)
```

We see that Chrysler Imperial, Fiat 128 and Toyota Corolla are the most outlying points and contribute to differ from normality in upper tail. The fact that the sample is not large (32 cars) does not help to check normality with Q-Q plot.

Conclusions

We have fitted several linear models to data for different cars and have choosen the one that presents a better compromise between data good description and low number of parameters. We see that the relevant magnitudes for fuel efficiency are weight, 1/4 mille time and transmission, being all of them significant. We conclude that manual transmission cars are more efficient.

Appendix 1: Results of fitting the model to all magnitudes

We perform a regression to all magnitudes to identify relevan parameters

```
lm1<-lm(mpg~.,mtcars)
summary(lm1)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Appendix 2: Model selection

We have used R step function to choose a model that represents well the data without having too much parameters.

```
AIC<-step(lm1,k=log(length(mtcars$mpg)))
```

```
summary(AIC)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, 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 ***
## am            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
```

We see that the best model is the one that relates mpg with wt (weight) qsec (1/4 mille time) and am (transmission). All parameters are relevant within a 5% confidence level.

Appendix 3: Residual Analysis

```r
lm3<-lm(mpg~wt+qsec+am,mtcars)
par(mfrow=c(2,2))
plot(lm3)
```