EXECUTIVE SUMMARY

This report aims at investigating the following questions on the base of the database mtcars:

1. Is an automatic or manual transmission better for MPG (miles per gallons)

2. Quantify the MPG difference between automatic and manual transmissions

The database contains 32 car models of which 19 have automatic transmission and 13 have manual.

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4          21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag      21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710         22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

First we check the null hypothesis that manual and automatic transmissions cars are from the same population (the variable "am" is 0 for automatic transmission and 1 for manual.). For the sake of this projects we assume "mpg" to have a normal distribution, although it looks a little skewed.

A simple t-test confirm that the two transmissions are not from the same population as the interval does not contain 0 and the p-value 0.0013736 is sufficiently low.

```
## [1] -11.280194  -3.209684
## attr(,"conf.level")
## [1] 0.95
```

We consider a linear model with "am" as a predictor and "mpg" as a response.
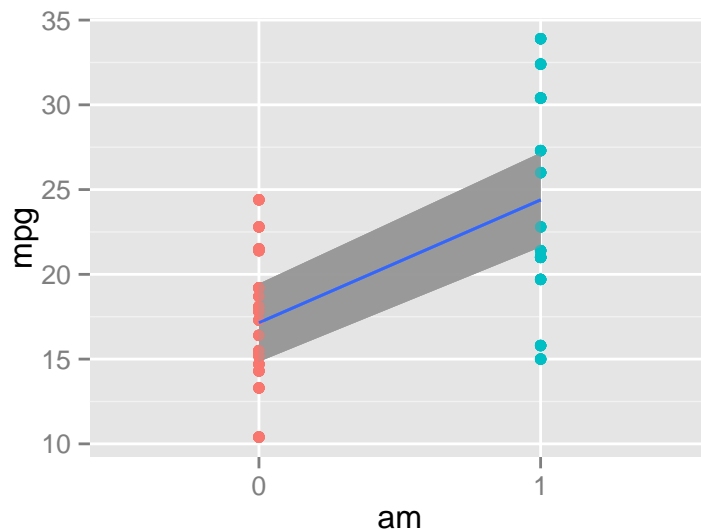

```
##
## Call:
## lm(formula = mpg ~ am - 1, data = mtcars)
##
```
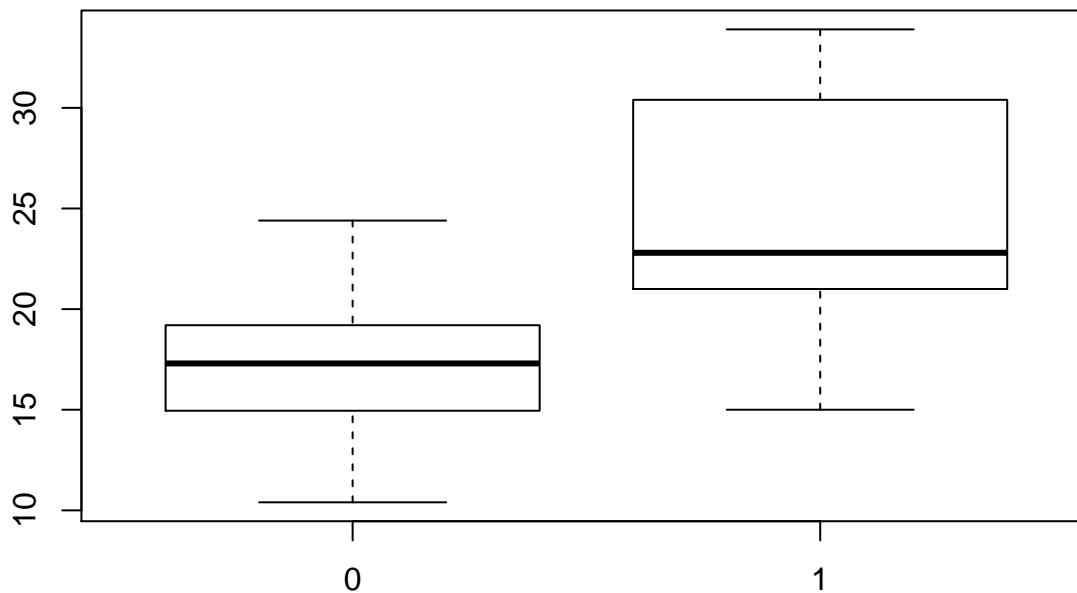
```
## Residuals:
##     Min     1Q  Median     3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##     Estimate Std. Error t value Pr(>|t|)
## am0   17.147      1.125   15.25 1.13e-15 ***
## am1   24.392      1.360   17.94  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9452
## F-statistic: 277.2 on 2 and 30 DF,  p-value: < 2.2e-16
```

The relatevly high F-statistic and the very low p-value confirm that there might be a correlation, but if we look at how well the model fit, R-squared 0.949 is very low (1 being best fit, 0 being no fit at all). This, together with a Residual Standard Error (RSE) of 4.822, suggests that the model is not very accurate and that there might be other model to better explain the correlation. We look therefore to other variables that might be confounder.

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + drat + wt + qsec +
##      vs + gear + carb - 1, data = mtcars)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## am0    23.87913   20.06582   1.190   0.2525
## am1    25.09125   18.63219   1.347   0.1981
## cyl6   -2.64870    3.04089  -0.871   0.3975
## cyl8   -0.33616    7.15954  -0.047   0.9632
## disp    0.03555    0.03190   1.114   0.2827
## hp     -0.07051    0.03943  -1.788   0.0939 .
## drat    1.18283    2.48348   0.476   0.6407
## wt     -4.52978    2.53875  -1.784   0.0946 .
## qsec    0.36784    0.93540   0.393   0.6997
## vs1     1.93085    2.87126   0.672   0.5115
## gear4   1.11435    3.79952   0.293   0.7733
## gear5   2.52840    3.73636   0.677   0.5089
## carb2  -0.97935    2.31797  -0.423   0.6787
## carb3   2.99964    4.29355   0.699   0.4955
## carb4   1.09142    4.44962   0.245   0.8096
## carb6   4.47757    6.38406   0.701   0.4938
## carb8   7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9817
```

```
## F-statistic:    102 on 17 and 15 DF,  p-value: 1.979e-12
```

The most influential confounder variable seems to be "wt", the weight.


```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
## am1         -0.02362    1.54565  -0.015    0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09


##
## Call:
## lm(formula = mpg ~ am * wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6004 -1.5446 -0.5325  0.9012  6.0909
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   31.4161      3.0201   10.402 4.00e-11 ***
## am1            14.8784      4.2640    3.489  0.00162 **
## wt             -3.7859      0.7856   -4.819 4.55e-05 ***
## am1:wt         -5.2984      1.4447   -3.667  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.591 on 28 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8151
## F-statistic: 46.57 on 3 and 28 DF,  p-value: 5.209e-11
```

APPENDIX Exploratory graphs