Coursera Regression Models Course Peer Reviewed Project
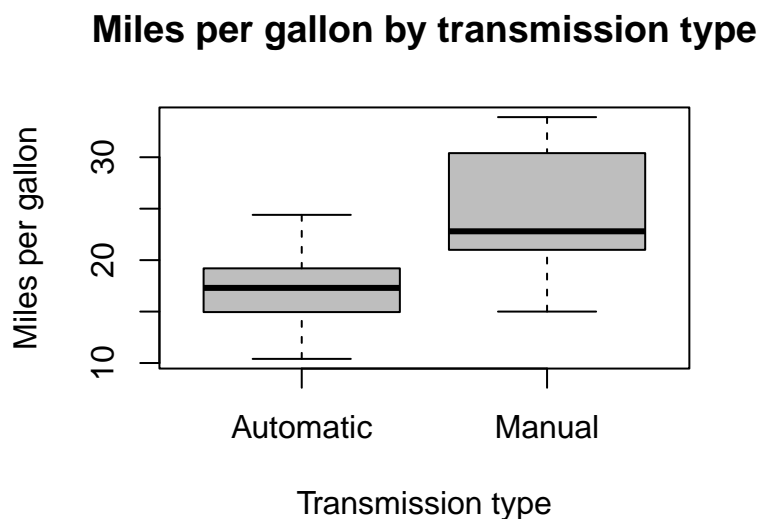
Sebastián ZL

EXECUTIVE SUMMARY

This project aims to explore some features that affect fuel consumption in miles per gallon (MPG) based on a collection of cars (mtcars - Motor Trend Car Road Tests) data set. This because Motor Trend Magazine is interested in exploring the relationship between a set of variables and miles per gallon (MPG). They are particularly interested in the following two questions: - Is an automatic or manual transmission better for MPG? - Quantifying how different is the MPG between automatic and manual transmissions?

EXPLORATORY ANALYSIS

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```


ANALYSIS

We proceed to see if theres a significant difference between average miles per galon beetwen the transmission type.

As we saw above, for a conventional significance level $\alpha = 0.05$ under the null hypothesis of true diffence in means equal to zero, we **reject** this hypothesis with a p-value of 0.0014. This is also confirmed by the 95% confidence interval of the means difference [-11.28,-3.20], which is not cointaining zero or near zero values. Now whe can proceed to calculate how different are the autonmyes.

```
##
## Call:
## lm(formula = mtcars$mpg ~ as.factor(mtcars$am))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             17.147      1.125  15.247 1.13e-15 ***
## as.factor(mtcars$am)1    7.245      1.764   4.106 0.000285 ***
## ---
## 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:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

The regression coefficient *"as.factor(mtcars$am)1"*; $\beta_{am} = 7.24$ tell us that **manual cars have an 7.245mpg average above the automatica ones.** *** indicates that its p-value its contaiden in *(0,001]* and its a statistically significant predictor. Unfurnately, as $R^2 = 0.3598$, only the 36% of the mpg variance its explained by the univariate model. Lets try another model including more predictors and their influence on it.

```
##
## Call:
## lm(formula = mpg ~ cyl + wt + as.factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    39.4179     2.6415  14.923 7.42e-15 ***
## cyl            -1.5102     0.4223  -3.576  0.00129 **
## wt             -3.1251     0.9109  -3.431  0.00189 **
## as.factor(am)1  0.1765     1.3045   0.135  0.89334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

With a multivariate model, the residual variation its decreased as evidenced in the $R^2 = 0.8303$ obtained. Also note by the significance levels, that the *gpm* its more stronghtly dependant in *cyl* and *wt* rather than in *am*. As well, *am* seems not to be independat variable as its coeficient changes with the presence of ther other variables to $\beta_{am} = 0.17$. A variance and resisual analysis can be found the Appendix.

APPENDIX

```
## [1] "A. VARIANCE ANALYSIS"


## Analysis of Variance Table
##
## Response: mpg
##               Df Sum Sq Mean Sq  F value    Pr(>F)
## cyl            1 817.71  817.71 119.8446 1.258e-11 ***
## wt             1 117.16  117.16  17.1714 0.0002854 ***
## as.factor(am)  1   0.12    0.12   0.0183 0.8933421
## Residuals     28 191.05    6.82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


## [1] "B. RESIDUAL PLOTS"
```
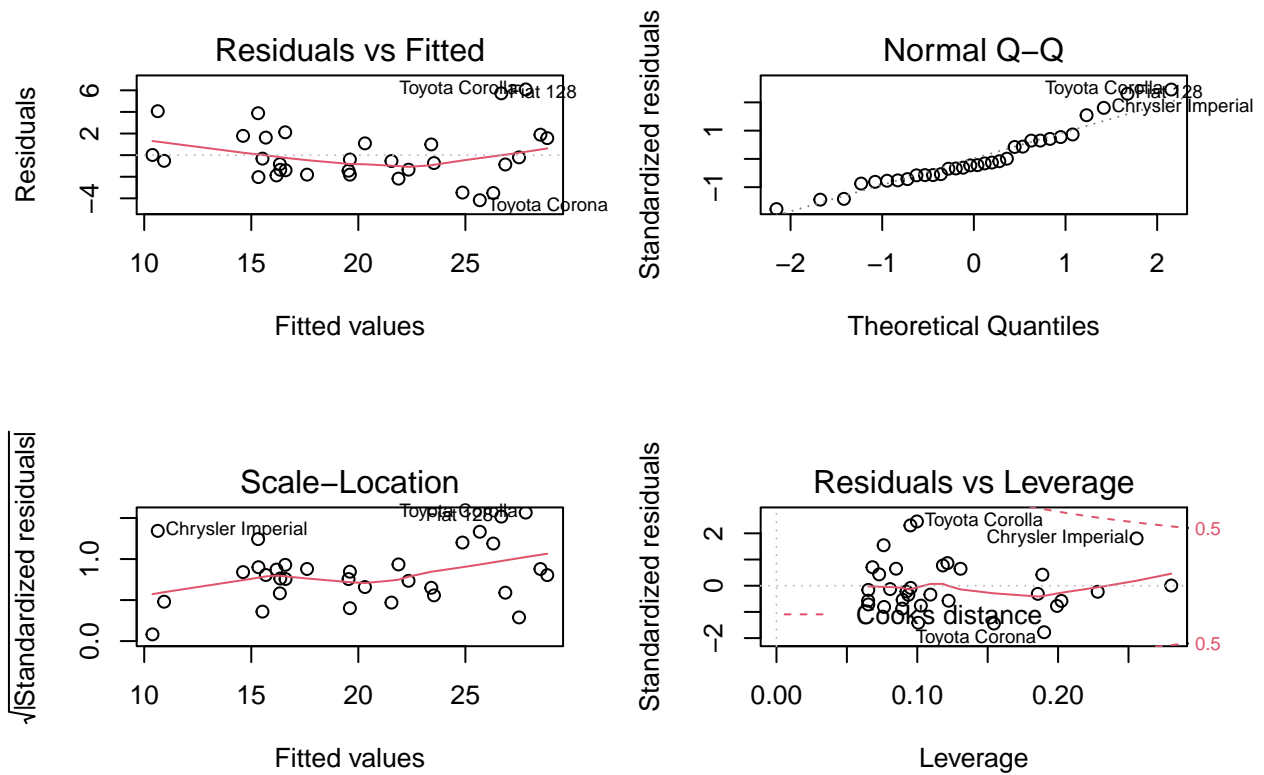


```
## [1] "C.DATASET SCATTERPLOT DATASET "
```