“Is an automatic or manual transmission better for MPG”

Motor Trend Analysis Executive Summary

For this project, we were asked the following: You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions: 1. “Is an automatic or manual transmission better for MPG” 2. “Quantifying how different is the MPG between automatic and manual transmissions?” I present below the steps I took to conduct this study. Load the data

```
data(mtcars)
str(mtcars)
```

A quick look at the data from str(mtcars) and ?mtcars indicates that some variables need to be changed as factor. Those variables are cyl, vs, gear, carb, and am.

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

Exploratory analysis

```
summary(mtcars)
```

See scatterplot matrix in the appendix. Regression model

I performed stepwise model selection using backwards elimination to determine the variables for the best model because it has the advantage to fit multiple models and find the best final model.

```
# backward elimination
full.model <- lm(mpg ~ ., data = mtcars)
best.model <- step(full.model, direction = "backward")

summary(best.model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Interpretation: The above procedure determines that the best model includes the cyl6, cyl8, hp, wt, and amManual variables (overall p-value<0.001). The adjusted R-squared indicates that about 84% of the variance is explained by the final model. Moreover, the output of this model suggests that mpg decreases with respect to cylinders (-3.03 and -2.16 for cyl6 and cyl8, respectively), horsepower (-0.03), and weight (for every 1,000lb, by -2.5). On the other hand, mpg increases with respect to having a manual transmission (by 1.8). Residual plots (see appendix) suggest that some transformation may be necessary to achieve linearity. Transmission type differences

I constructed a boxplot of mpg per transmission type (see appendix) and conducted a t-test as follows:

```
t.test(mpg ~ am, data = mtcars) # results not shown because of page
limitation

##
##  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
```

Interpretation: The boxplots show a difference in mpg depending on the type of transmission. The t-test output confirms that this difference is statistically significant (p-value < 0.05). Conclusions

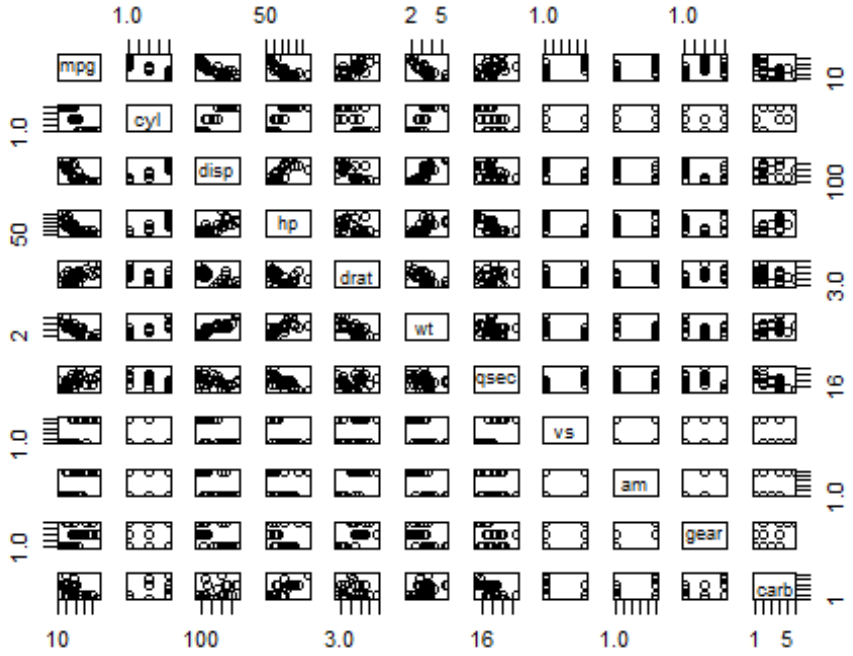
According to these results, cars with a manual transmission are better for mpg than cars with an automatic transmission. The rate of change of the conditional mean mpg with respect to am is about 1.8, and we are 95% confident that this value varies between -1.06 and 4.68. There are however some limitations to this study. To name a few:

```
Conducting this study with the base package only makes it difficult to dig
deeper into this assignment.
A lack of linearity in the residual plots. This could have been adressed by
transforming the variables in an attempt to achieve linearity and would have
been facilitated by the use of packages other than the base to determine
which transformations are necessary.
The sample size is very small, with is a limitation by itself for statitical
inference.
Being allowed only 5 pages (including 2 pages or less for the main text) to
conduct this study.
```

```
Appendix : Supporting figures
```

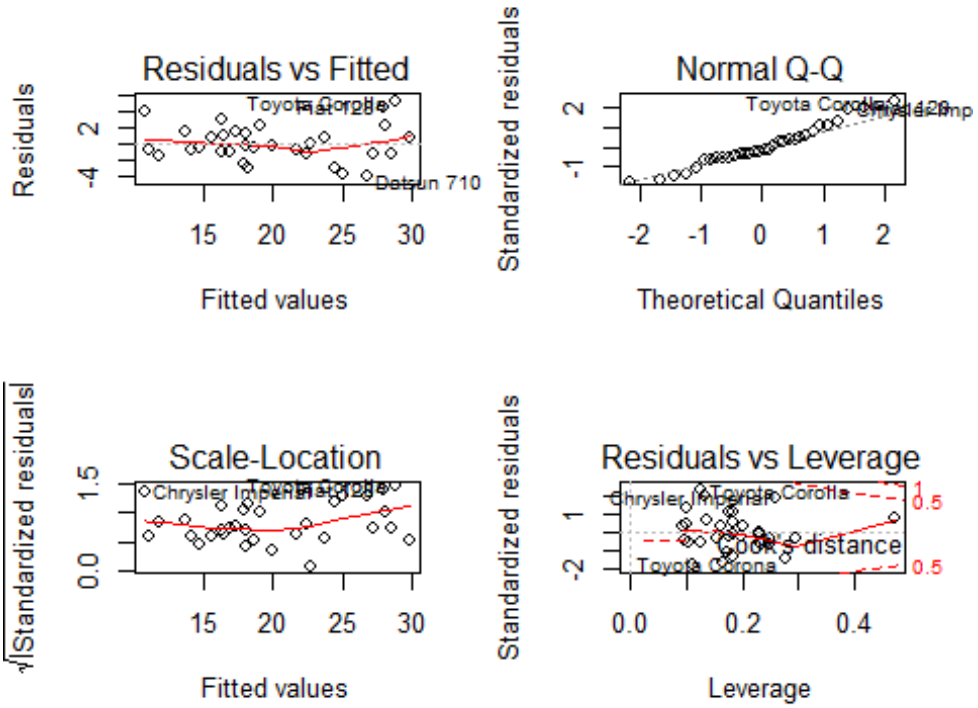
Scatterplot matrix of the “mtcars” dataset

```
pairs(mpg ~ ., data = mtcars)
```



Residual plots of the best model as evaluated by the stepwise regression

```
par(mfrow=c(2, 2))
plot(best.model)
```



Boxplot of miles per gallon by transmission type

```
boxplot(mpg ~ am, data = mtcars, col = "blue", ylab = "miles per gallon")
```