

# Regression Models Project

## 1. Introduction

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:

- “Is an automatic or manual transmission better for MPG”
- “Quantify the MPG difference between automatic and manual transmissions”

## 2. Getting data

```
library(datasets)
data(mtcars)
head(mtcars)
```

```
##           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
```

```
summary(mtcars)
```

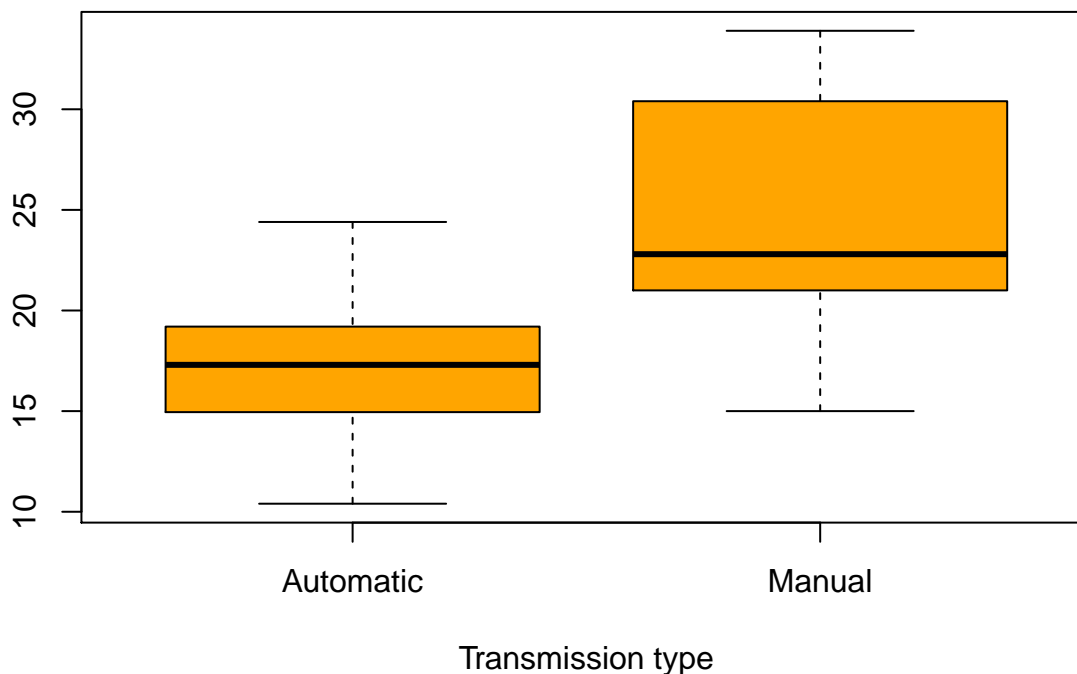
```
##           mpg           cyl           disp           hp
##  Min.   :10.40  Min.   :4.000  Min.   : 71.1  Min.   : 52.0
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5
## Median :19.20 Median :6.000 Median :196.3 Median :123.0
## Mean   :20.09 Mean   :6.188 Mean   :230.7 Mean   :146.7
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0
## Max.   :33.90 Max.   :8.000 Max.   :472.0 Max.   :335.0
##           drat           wt           qsec           vs
##  Min.   :2.760  Min.   :1.513  Min.   :14.50  Min.   :0.0000
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000
## Mean   :3.597 Mean   :3.217 Mean   :17.85 Mean   :0.4375
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000
## Max.   :4.930 Max.   :5.424 Max.   :22.90 Max.   :1.0000
##           am           gear           carb
##  Min.   :0.0000  Min.   :3.000  Min.   :1.000
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000
## Median :0.0000 Median :4.000 Median :2.000
## Mean   :0.4062 Mean   :3.688 Mean   :2.812
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000
## Max.   :1.0000 Max.   :5.000 Max.   :8.000
```

### 3. Regression Analysis

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

fullmodel <- lm(mpg ~ am, data = mtcars)
bestmodel <- lm(mpg ~ cyl + hp + wt + am, data = mtcars)

boxplot(mpg ~ am,
        data = mtcars,
        col = "orange",
        xlab = "Transmission type")
```



```
summary(fullmodel)
```

```
##
## Call:
## lm(formula = mpg ~ am, 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|)
## (Intercept)    17.147     1.125   15.247 1.13e-15 ***
## amManual         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
```

```
summary(bestmodel)
```

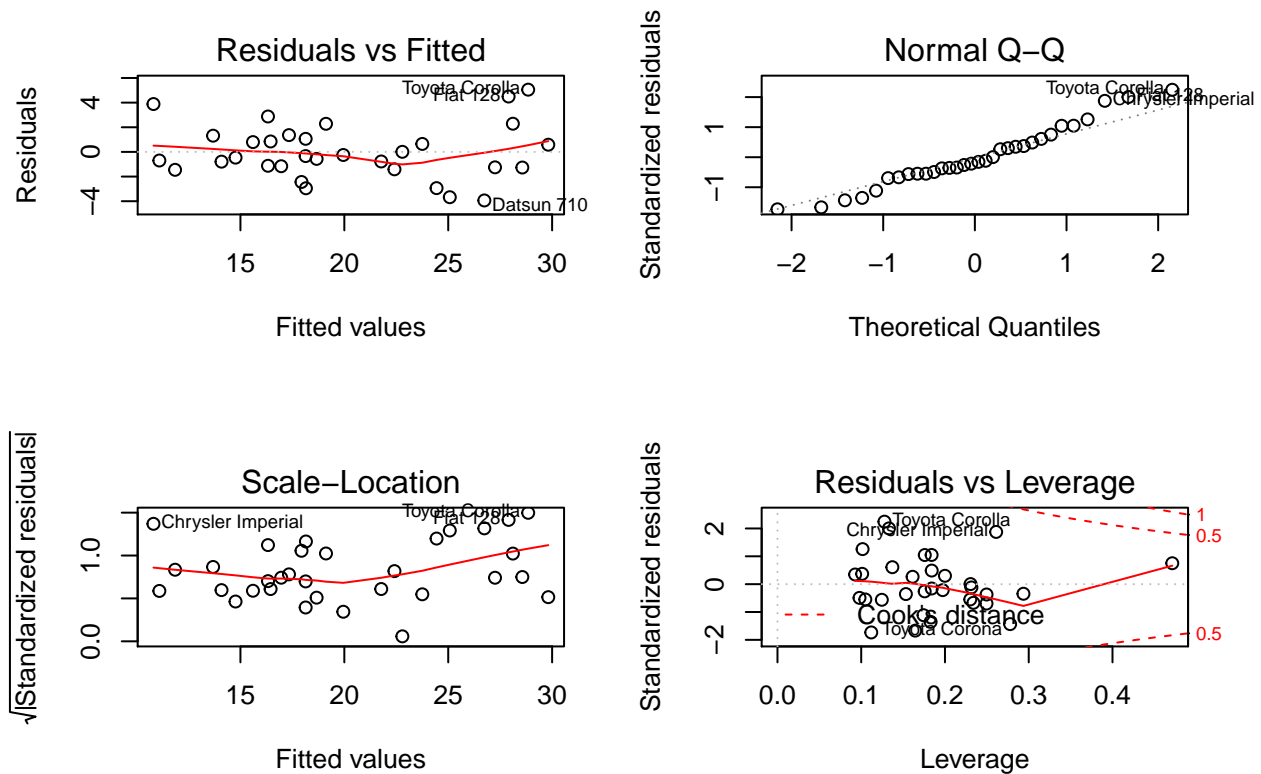
```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.70832    2.60489   12.940 7.73e-13 ***
## 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
```

We are looking at the “Adjust r-squared” value of both models can be seen as the bestmodel obtains a coefficient of 0.8401, higher than the full\_model: 0.3385. We can say That approximately 84% of the variability is Explained by the combined model above.

## 4. Model graph

Additionally, we also plot the residuals to examine any heteroskedacity between the fitted and residual values; as well as to check for any non-normality.

```
par(mfrow = c(2, 2))
plot(bestmodel)
```



From the above graphs, we can make a few observations about the combined model. The randomness of the distribution of the points in the Residuals vs. Fitted graph confirms the variable independence. The linearity of the Normal Q-Q graph indicates that the residuals are distributed under a normal distribution. The labeled points appear to be leverage points above the rest of the points.

## 5. Statistical Inference

Here we perform a t-test on the different transmissions versus mpg.

```
t.test(mpg ~ am, data = mtcars)

##
## 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
```

From these results, we can reject the null hypothesis saying that the effect on mpg of manual and automatic transmissions are the same.

## 6. Conclusions

Looking at the combined model, we can see how mpg is effected by changes in cyl, hp and wt. Cars with manual transmission get about 1.8 MPG more than automatic transmission. MPG decreases by about 2.5 for every 1000 pound increase in weight. MPG decreases very marginally with horsepower, about 3 MPG for every 100 horsepower.