Regression Models Course Project

Armulfo Perez

12/5/2020

Executive summary

We 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."

Based on the mcars dataset, the most important factor for mileage is the weight of the vehicle. Manual transmission is more efficient for light vehicles, but it gets less important as vehicles become heavier. Manual transmission is more efficient than automatic for lighter cars (as much as 30%), but as car become heavier, the difference reverses.

Data

The mtcars data set consists of 32 observations of 11 variables, including miles per gallon performance, and whether a car has automatic or manual transmission.

```
str(mtcars)
```

```
## '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 ...
```

Basic Analysis

Let´s start by determining the relative importance of the transmission type on millages performance by fitting a linear model to the data.

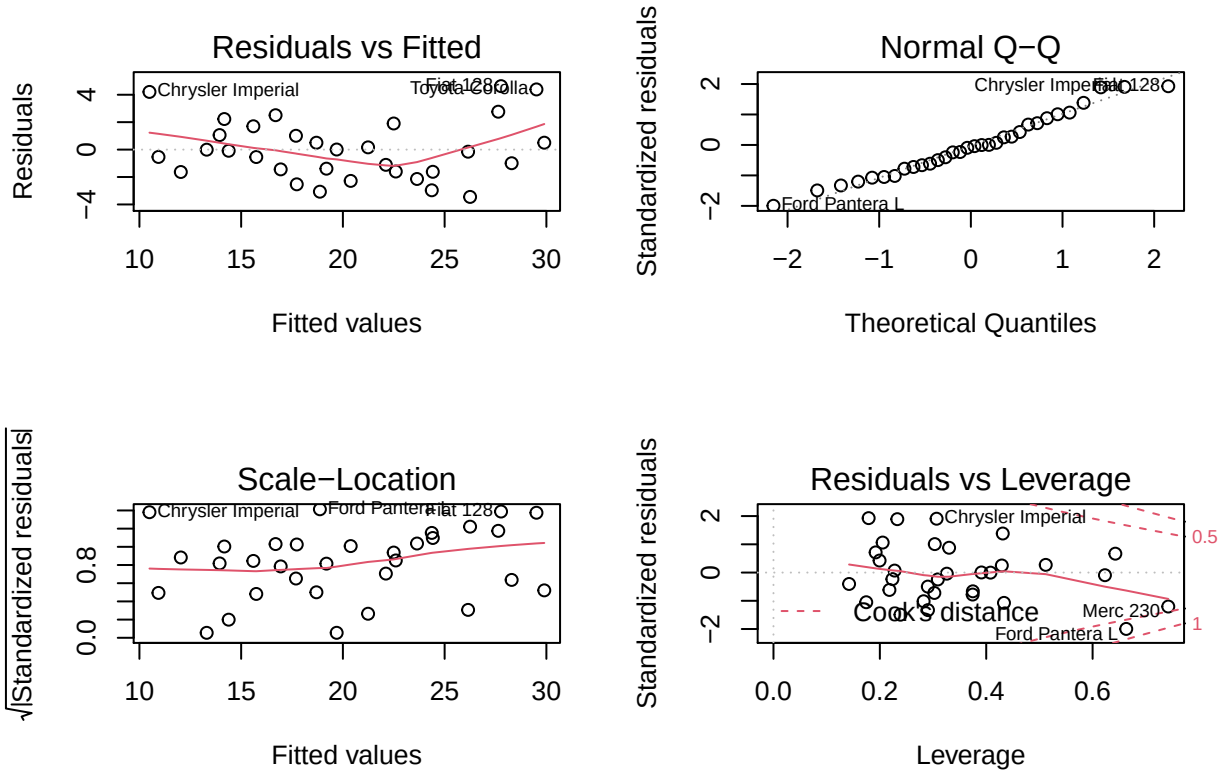
```
lmall <- lm(mpg ~ ., data = mtcars)
lmall
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Coefficients:
```

```
## (Intercept)         cyl        disp          hp        drat          wt
##    12.30337    -0.11144     0.01334    -0.02148     0.78711    -3.71530
##        qsec          vs          am        gear        carb
##     0.82104     0.31776     2.52023     0.65541    -0.19942
```

The model shows that the most important factor determining millages is the weight of the vehicle, followed by transmission type, with a coefficient of 2.5.

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




```
dev.off()
```

```
## null device
##           1
```

The residuals plots show that the normalcy condition is meet and that there are points with high leverage.

Weight and transmission type

Let's consider the relation between millage and vehicle weight under the two types of transmission.

```
mcars <- mtcars[mtcars$am==1,]
acars <- mtcars[mtcars$am==0,]
summary(mcars)[,c(1,6)]
```

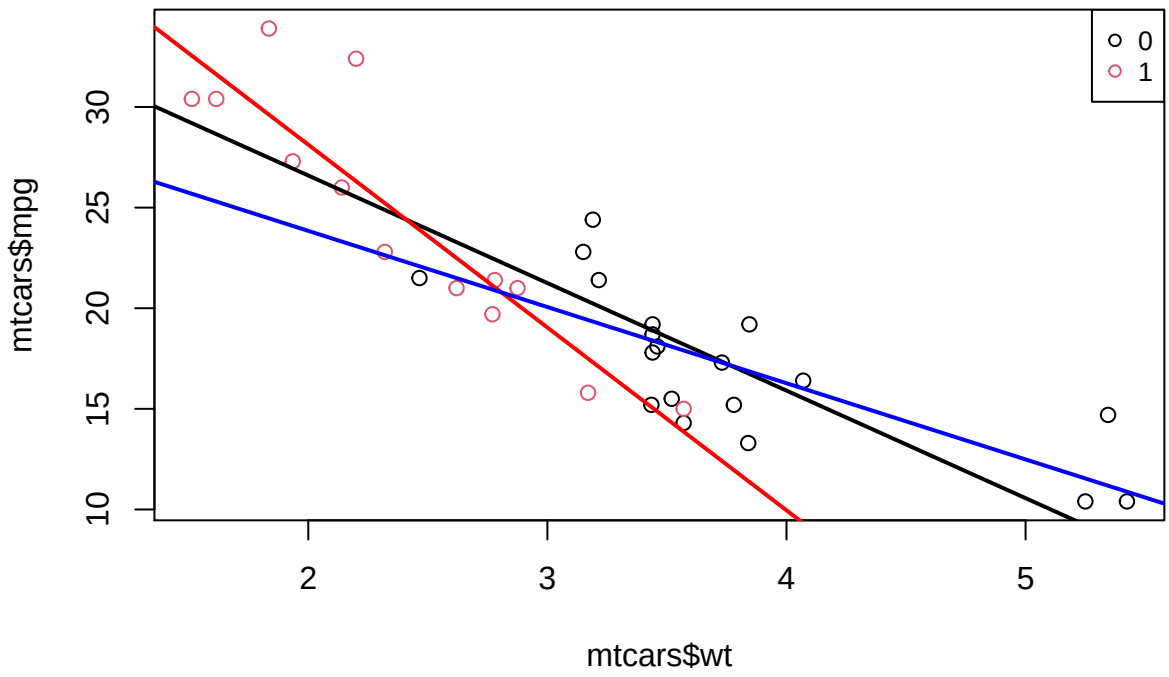
```
##       mpg             wt
##  Min.   :15.00   Min.   :1.513
##  1st Qu.:21.00   1st Qu.:1.935
```

```
##  Median :22.80   Median :2.320
##  Mean   :24.39   Mean   :2.411
##  3rd Qu.:30.40   3rd Qu.:2.780
##  Max.   :33.90   Max.   :3.570
```

```
summary(acars)[,c(1,6)]
```

```
##       mpg             wt
##  Min.   :10.40   Min.   :2.465
##  1st Qu.:14.95   1st Qu.:3.438
##  Median :17.30   Median :3.520
##  Mean   :17.15   Mean   :3.769
##  3rd Qu.:19.20   3rd Qu.:3.842
##  Max.   :24.40   Max.   :5.424
```

```
lmt <- lm(mpg ~ wt, data = mtcars)
lma <- lm(mpg ~ wt, data = acars)
lmm <- lm(mpg ~ wt, data = mcars)
plot(mtcars$wt,mtcars$mpg,col=mtcars$am+1)
legend('topright', legend = levels(factor(mtcars$am)), col = 1:2, cex = 0.8, pch = 1)
abline(lmt,col="black",lw=2)
abline(lmm,col="red",lw=2)
abline(lma,col="blue",lw=2)
```



The plot shows the automatic transmission cars as black dots, and manual transmission one's as red. Three linear models are fitted (black: all cars, blue: automatic, red: manual). To more clearly see the effect of manual transmission, let's eliminate extreme cases, those weighting more than 5 tons and with better mileage than 31 mpg.

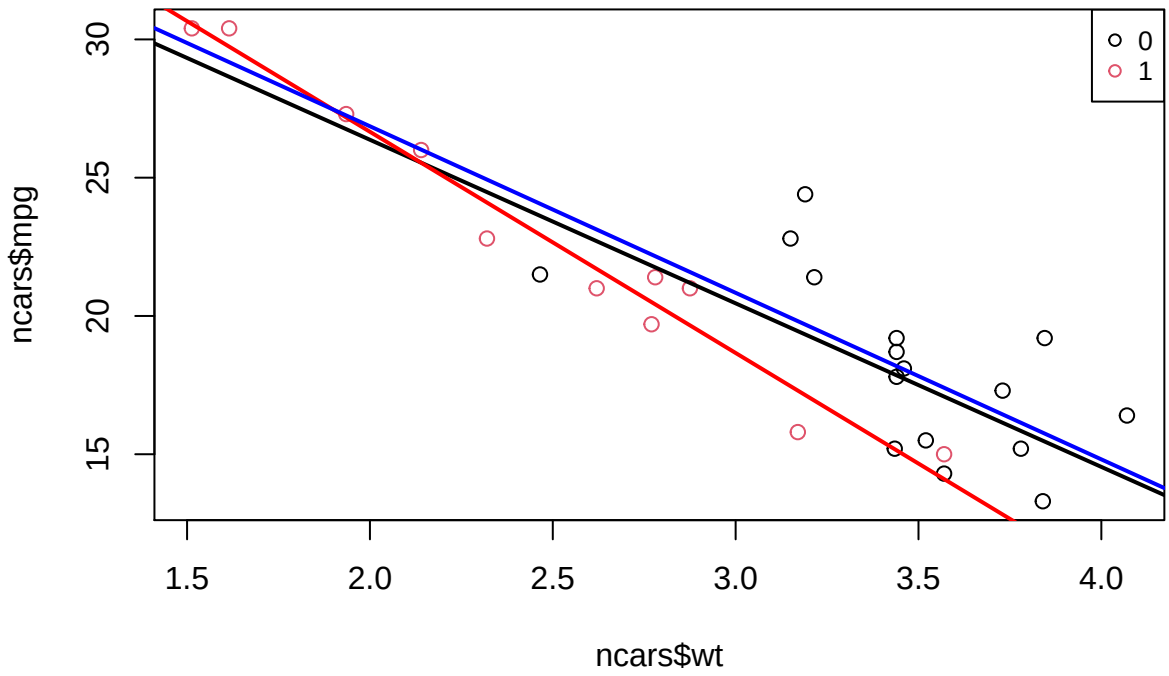
```r
ncars <- mtcars[mtcars$mpg < 31 & mtcars$wt<5,]
mncars <- ncars[ncars$am==1,]
ancars <- ncars[mtcars$am==0,]
lmn <- lm(mpg ~ wt, data = ncars)
lmn
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = ncars)
##
## Coefficients:
## (Intercept)           wt
##      38.201       -5.915
```

```r
predict.lm(lmn, interval="confidence")
```

```
##                       fit      lwr      upr
## Mazda RX4         22.70418 21.64155 23.76681
## Mazda RX4 Wag     21.19589 20.27295 22.11883
## Datsun 710        24.47864 23.16105 25.79622
## Hornet 4 Drive    19.18484 18.27773 20.09195
## Hornet Sportabout 17.85400 16.84511 18.86288
## Valiant           17.73570 16.71421 18.75719
## Duster 360        17.08506 15.98585 18.18428
## Merc 240D         19.33271 18.43176 20.23366
## Merc 230          19.56930 18.67585 20.46276
## Merc 280          17.85400 16.84511 18.86288
## Merc 280C         17.85400 16.84511 18.86288
## Merc 450SE        14.12764 12.55557 15.69970
## Merc 450SL        16.13869 14.90548 17.37190
## Merc 450SLC       15.84295 14.56382 17.12207
## Honda Civic       28.64861 26.56155 30.73567
## Toyota Corona     23.62098 22.43572 24.80625
## Dodge Challenger  17.38081 16.31860 18.44302
## AMC Javelin       17.88357 16.87776 18.88938
## Camaro Z28        15.48805 14.15174 16.82436
## Pontiac Firebird  15.45848 14.11731 16.79965
## Fiat X1-9         26.75586 25.03565 28.47606
## Porsche 914-2     25.54331 24.04475 27.04187
## Lotus Europa      29.25193 27.04432 31.45953
## Ford Pantera L    19.45101 18.55417 20.34784
## Ferrari Dino      21.81695 20.84767 22.78623
## Maserati Bora     17.08506 15.98585 18.18428
## Volvo 142E        21.75780 20.79367 22.72193
```

```r
lmna <- lm(mpg ~ wt, data = ancars)
lmnm <- lm(mpg ~ wt, data = mncars)
plot(ncars$wt,ncars$mpg,col=ncars$am+1)
legend('topright', legend = levels(factor(mtcars$am)), col = 1:2, cex = 0.8, pch = 1)
abline(lmn,lw=2)
abline(lmnm,col="red",lw=2)
abline(lmna,col="blue",lw=2)
```



It is evident from the plot that manual transmission is more efficient than automatic for lighter cars (as much as 30%), but as car become heavier, the difference reverses. According to the prediction confidence interval estimates of the model the predicted milages is estimated roughly with +/- 2 mpg.

Conclusions.

The most important factor for mileage is the weight of the vehicle. Manual transmission is more efficient for light vehicles but it gets less important as vehicles become heavier. According to the prediction confidence interval estimates of the model the predicted milages is estimated roughly with +/- 2 mpg.