Regression Modeling Course Project

Yigit Ozan Berk

8/16/2019

Summary: This analysis tries to find answers to following questions:

1 - Is an automatic or manual transmission better for MPG? 2 - Quantify the MPG difference between automatic and manual transmissions

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Initializing

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

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

Our dataset consists of 11 variables:

[, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs Engine (0 = V-shaped, 1 = straight) [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors

```
mtcars$am <- as.factor(mtcars$am)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$cyl <- as.factor(mtcars$cyl)
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
```

Exploratory Data Analysis

Checking out if the automatic and manual transmission groups have different mean miles per gallon

```
auto <- mtcars %>% filter(am == "0" )
manual <-mtcars %>% filter(am == "1")
summary(auto$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   10.40   14.95   17.30   17.15   19.20   24.40
```

```
summary(manual$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   15.00   21.00   22.80   24.39   30.40   33.90
```

T-Test

```
t.test(manual$mpg, auto$mpg, paired = F, var.equal = F)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  manual$mpg and auto$mpg
## 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:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y 
##  24.39231  17.14737
```

Our T-test gives that the true difference in means miles per gallon between manual and automatic transmission is not equal to 0.

Linear Regression

Initial model with one variable:

```
fit1 = lm(mpg ~ am, data= mtcars)
summary(fit1)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

The transmission effect seems significant at first glance. According to our initial model, it seems that switching from automatic to manual transmission increases the range of the car by 7.24 miles per gallon. But we need to analyze for possible confounders.

The most likely possible confounders that may affect the mpg are weight and hp. We also need to check for others in general.

```
fit1.1 = lm(mpg ~ am + hp, data = mtcars)
summary(fit1.1)$coef
```

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
```

```
## am1          5.2770853 1.079540576  4.888270 3.460318e-05
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
```

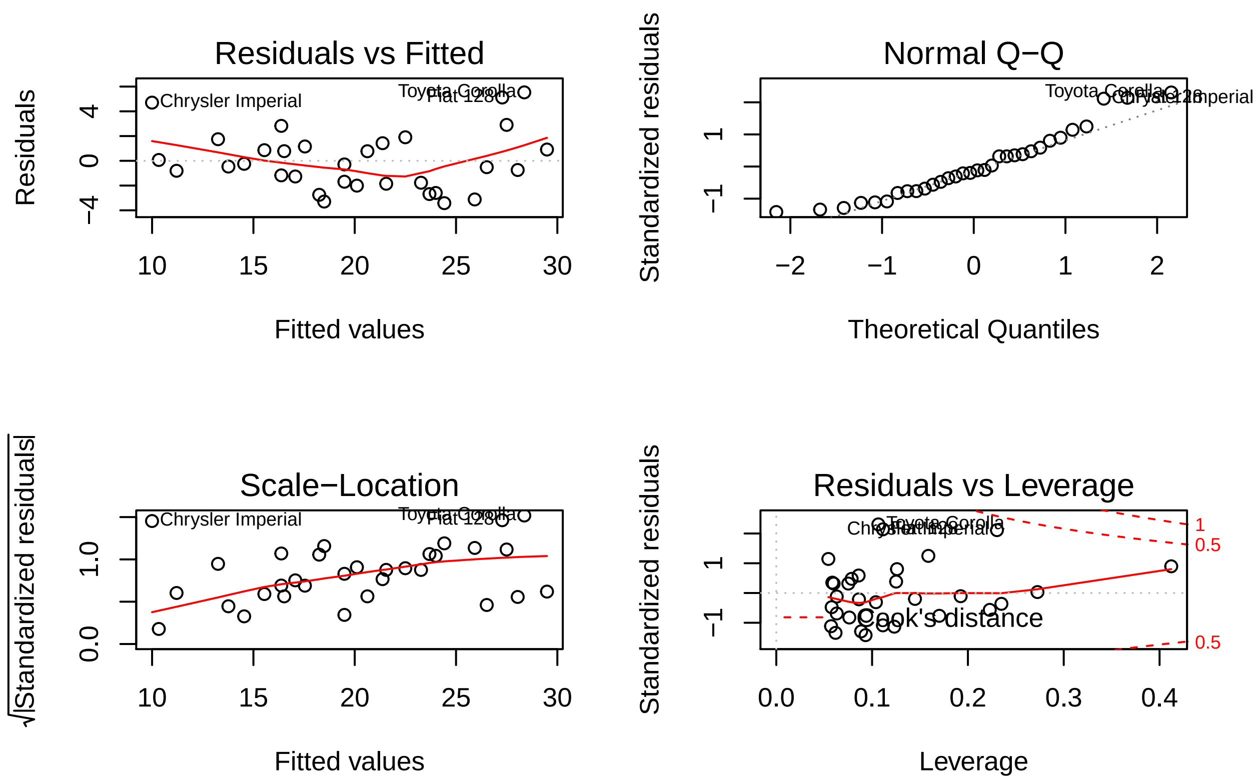
```
fit1.2 = lm(mpg ~ am + hp + wt, data = mtcars)
fit1.3 = lm(mpg ~ am + hp + wt + cyl, data = mtcars)
anova(fit1, fit1.1, fit1.2, fit1.3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + cyl
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 81.8529 1.634e-09 ***
## 3     28 180.29  1     65.15 11.2157  0.002484 **
## 4     26 151.03  2     29.27  2.5191  0.099998 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

comparing the nested models, it seems the cylinder variable is redundant. However, hp and wt variables influence the residual variance significantly.

Let's check the residuals for patterns for the hp and wt variables

```
par(mfrow = c(2,2))
plot(fit1.2)
```



It seems there is some non-linearity. We solve this problem by modeling the interaction between hp and wt variables.

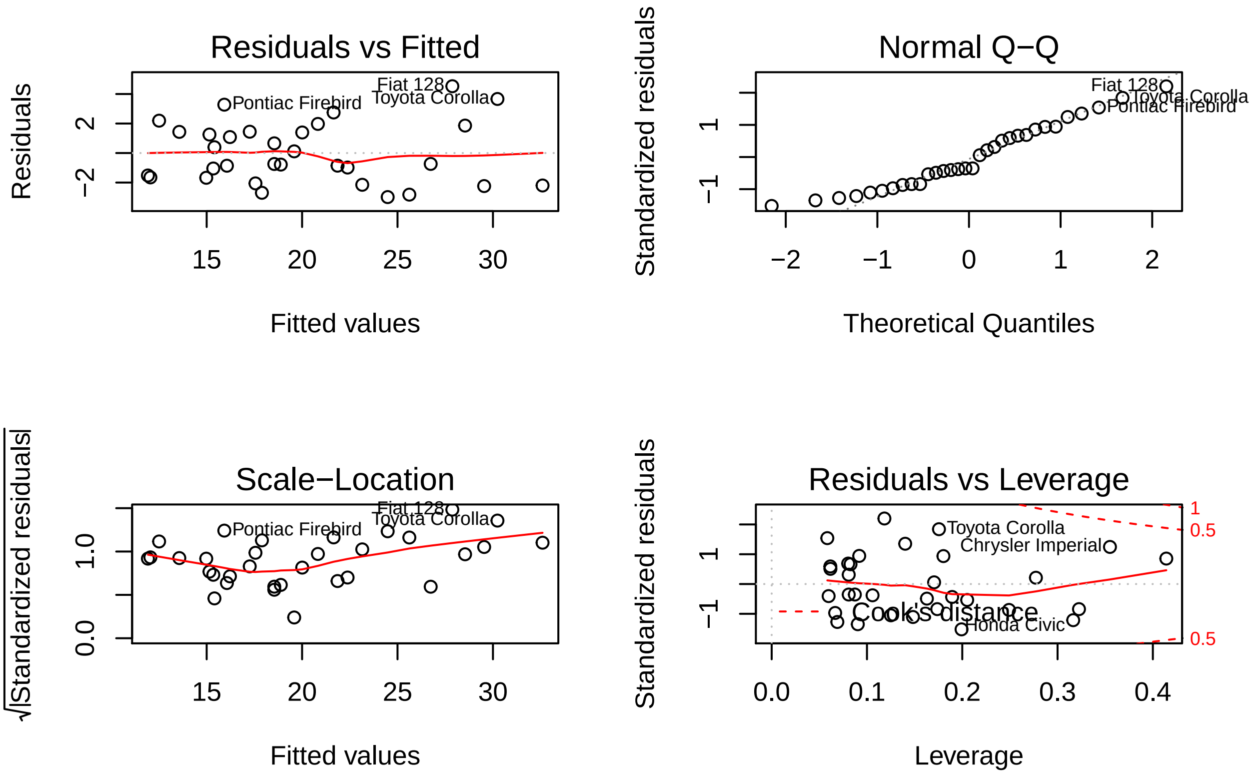
```r
fit1.4 = lm(mpg ~ am + hp*wt, data = mtcars)
summary(fit1.4)
```

```
##
## Call:
## lm(formula = mpg ~ am + hp * wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9845 -1.6580 -0.7407  1.4362  4.5266
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.452241   5.280731   9.365 5.69e-10 ***
## am1          0.125107   1.333431   0.094 0.925942
## hp          -0.119303   0.026550  -4.494 0.000119 ***
## wt          -8.100558   1.789325  -4.527 0.000108 ***
## hp:wt        0.027488   0.008473   3.244 0.003130 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.192 on 27 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8677
## F-statistic: 51.84 on 4 and 27 DF,  p-value: 2.765e-12
```

```r
confint(fit1.4)
```

```
##                   2.5 %      97.5 %
## (Intercept)  38.61707633 60.28740526
## am1          -2.61086742  2.86108128
## hp           -0.17377926 -0.06482709
## wt          -11.77194963 -4.42916547
## hp:wt         0.01010407  0.04487245
```

```r
par(mfrow = c(2,2))
plot(fit1.4)
```



Our final model consists of the coefficients of am, hp, wt and the interaction between hp:wt.

```
dfbetas(fit1.4)
```

```
##                      (Intercept)          am1           hp           wt
## Mazda RX4             0.2229937740 -0.312719480 -0.149936156 -0.239169899
## Mazda RX4 Wag         0.1445912590 -0.177327541 -0.089536903 -0.153744133
## Datsun 710            0.0630466614 -0.243523618 -0.010845917 -0.091073466
## Hornet 4 Drive       -0.0053010247 -0.074748765  0.045559204  0.034426502
## Hornet Sportabout     0.0270406849 -0.116376753  0.057621386 -0.022827417
## Valiant               0.0473906064  0.002779453 -0.048100717 -0.067184891
## Duster 360           -0.1157407236  0.175403312 -0.005083592  0.136720055
## Merc 240D            -0.0864521529 -0.054597639  0.062062560  0.193554047
## Merc 230              0.0022441693 -0.107574183  0.043748776  0.047893378
## Merc 280             -0.0234790509 -0.017029520  0.034699317  0.035858725
## Merc 280C             0.0265555950  0.019260959 -0.039246093 -0.040557423
## Merc 450SE           -0.0353741886 -0.010652726  0.031631024  0.044522459
## Merc 450SL           -0.0003565963 -0.053153837  0.034565466  0.004768639
## Merc 450SLC           0.0037927751  0.036999472 -0.026540786 -0.007693670
## Cadillac Fleetwood    0.0489783149 -0.143818559  0.175266217 -0.072042317
## Lincoln Continental  -0.0274541972 -0.125972309  0.281747982  0.006638099
## Chrysler Imperial     0.2547611816  0.042056009 -0.589309239 -0.239573418
## Fiat 128              0.1823813720  0.305243229 -0.341848505 -0.094910944
## Honda Civic          -0.6864918866  0.232760800  0.654746754  0.638257214
## Toyota Corolla        0.6014043045 -0.094015963 -0.590863169 -0.541211197
## Toyota Corona        -0.5579251360  0.667280715  0.235609688  0.494313253
## Dodge Challenger      0.0375862716  0.086967597 -0.100132448 -0.060883017
## AMC Javelin           0.0187420157  0.150042822 -0.125467945 -0.047542506
## Camaro Z28           -0.1937882058  0.238923121  0.061783424  0.221367785
## Pontiac Firebird     -0.0621371938 -0.097193846  0.117278850  0.084797810
```

```
## Fiat X1-9          -0.2679882518 -0.001644692  0.282719515  0.232420086
## Porsche 914-2      -0.0167904227 -0.035182315  0.018518867  0.010481086
## Lotus Europa        0.2553329897 -0.134207544 -0.088444398 -0.264284699
## Ford Pantera L     -0.0588654118  0.055208267  0.079139919  0.041264917
## Ferrari Dino       -0.0168494727  0.016736561  0.018008321  0.015371107
## Maserati Bora      -0.0094743721  0.110779351  0.126712083 -0.111118748
## Volvo 142E          0.1396981118 -0.178812613 -0.087625690 -0.149241813
##                          hp:wt
## Mazda RX4           0.20076249
## Mazda RX4 Wag       0.12237627
## Datsun 710          0.06123911
## Hornet 4 Drive     -0.05965579
## Hornet Sportabout  -0.02847516
## Valiant             0.06772809
## Duster 360         -0.06688015
## Merc 240D          -0.17617027
## Merc 230           -0.07611081
## Merc 280           -0.04273871
## Merc 280C           0.04833891
## Merc 450SE         -0.03340471
## Merc 450SL         -0.02198193
## Merc 450SLC         0.01821290
## Cadillac Fleetwood -0.11854453
## Lincoln Continental -0.22746583
## Chrysler Imperial   0.55340692
## Fiat 128            0.21164818
## Honda Civic        -0.64812939
## Toyota Corolla      0.55891525
## Toyota Corona      -0.29761868
## Dodge Challenger    0.09769849
## AMC Javelin         0.11247753
## Camaro Z28         -0.15687446
## Pontiac Firebird   -0.10373377
## Fiat X1-9          -0.25348568
## Porsche 914-2      -0.01137519
## Lotus Europa        0.15155877
## Ford Pantera L     -0.05689438
## Ferrari Dino       -0.01707888
## Maserati Bora       0.04867646
## Volvo 142E          0.12000908
```

From a quick glance at the dfbetas, we can see that Honda Civic, Toyota Corolla, and Cyrysler Imperial are top 3 influential observations. However, the data is clean and tidy. It is very unlikely that these observations are errors.

```
cooks.distance(fit1.4)
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##        0.0315169119        0.0088332478        0.0361539163
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##        0.0079471836        0.0082362307        0.0033871388
##          Duster 360           Merc 240D            Merc 230
##        0.0148082243        0.0596024980        0.0180460527
##            Merc 280           Merc 280C          Merc 450SE
##        0.0017206544        0.0021989009        0.0045753632
```

```
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood 
##        0.0034112694        0.0020847127        0.0501392462 
## Lincoln Continental   Chrysler Imperial            Fiat 128 
##        0.0676055149        0.1707166771        0.1298085785 
##         Honda Civic      Toyota Corolla       Toyota Corona 
##        0.1363949262        0.1442959822        0.1152196007 
##    Dodge Challenger         AMC Javelin          Camaro Z28 
##        0.0134717409        0.0240311659        0.0298254696 
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2 
##        0.0294347423        0.0426635950        0.0023794370 
##        Lotus Europa      Ford Pantera L        Ferrari Dino 
##        0.0386442240        0.0034148513        0.0001347569 
##       Maserati Bora          Volvo 142E 
##        0.1032192228        0.0093545089
```

With a final look at cooks.distance, we can see that there are no outliers in this dataset, creating bias in our model.

Conclusion

We can conclude that ;

- We need more data to see if transmission type makes a difference in mpg variable. With the current data set, we can only conclude with %7.41 confidence that the transmission type changes miles per gallon between the values -2.61086742 and 2.86108128. We cannot conclude if transmission type affects mpg positively or negatively. With the current data set, the statistics test do not provide a significant result(%92.59 > %5). Maybe other variables can be included such us driven kms in city/in suburbs to find out the effect of traffic.
- 1 point increase in HP results in 0.12 decrease in miles per gallon.
- 1000 pounds increase in weight results in 8.10 decrease in miles per gallon.
- the effect of hp on mpg somewhat depends on the weight of the car.