Regression Model Project

Mayank Pundhir

October 26, 2014

Summary

The following Project has been done for Motor Trend, a magazine about the automobile industry. Data was taken from 'mtcars' dataset from R which was extracted from the 1974 Motor Trend US magazine, and comprised fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). The magazine was interested in exploring the relationship between a set of variables and miles per gallon (MPG). They were particularly interested in the following two questions:

- Was an automatic or manual transmission better for MPG
- Quantify the MPG difference between automatic and manual transmissions

Manual transmission was found to be more fuel efficient than the other.
Model 1 calculated that manual is better than auto transmission by a difference of 2.93 mpg.
Model 2 suggested that the difference was 5.22.

Part 1 of the Project: Qualitative Approach

As transmission is given by am, it was first converted to factors: 'auto' and 'manual'. As fuel economy of a car depends on multidimensional aspects such as it's engine, size. We first plotted mpg against weight and horsepower. Mpg vs weight graph showed some ambiguous results whereas mpg vs horsepower showed a significant advantage of manual over automatic transmission. Automatic cars were in mid to heavy weight whereas manual cars were in light to mid weight range. Some low fuel economy manual cars were analyzed(Least 0.25 quantile).

```
man<-mtcars[mtcars$am=="manual",]
 man[man$mpg<quantile(man$mpg,0.25),]

##                mpg cyl disp  hp drat   wt qsec vs     am gear carb
## Ford Pantera L 15.8   8  351 264 4.22 3.17 14.5  0 manual    5    4
## Ferrari Dino   19.7   6  145 175 3.62 2.77 15.5  0 manual    5    6
## Maserati Bora  15.0   8  301 335 3.54 3.57 14.6  0 manual    5    8
```

Ford Pantera and Maserati Bora were very powerful and fast whereas Ferrari Dino was less horsepower but was almost as fast as the other two. These were not designed with fuel economy in mind. Similarly high fuel economy automatic acars were analyzed(Best 0.25 quantile).

```
auto<-mtcars[mtcars$am=="auto",]
 auto[auto$mpg>quantile(auto$mpg,0.75),]

##                mpg cyl  disp  hp drat    wt  qsec vs   am gear carb
## Hornet 4 Drive 21.4   6 258.0 110 3.08 3.215 19.44  1 auto    3    1
## Merc 240D      24.4   4 146.7  62 3.69 3.190 20.00  1 auto    4    2
## Merc 230       22.8   4 140.8  95 3.92 3.150 22.90  1 auto    4    2
## Toyota Corona  21.5   4 120.1  97 3.70 2.465 20.01  1 auto    3    1
```

All of the cars had small horsepower. They were slow but fuel efficient.
Both automatic and manual cars were trading fuel efficiency or power and speed for another. But as seen from the graph manual transmission was more fuel efficient than automatic transmission.

Part 2 of the Project: Quantitative analysis

Since many variables are highly correlated to each other so only few variables were selected.

```
fit1<-lm(mpg~hp+wt+qsec+am,mtcars)
fit2<-lm(mpg~hp*am,mtcars)
```

variance Inflation by different variables was calculated.

```
vif(fit1)

##    hp    wt  qsec    am
## 4.922 3.965 3.216 2.542

vif(fit2)

##    hp    am hp:am
## 2.786 6.255 6.460
```

Model 'fit1' was appropriate as it had 4 variables out of 10 but it did not consider the interaction of variables.

Model 'fit2' was concise and also addressed the interaction term though it was negligible 4.0289×10^{-4}.Variance Inflation by 'am' in Model 2 was more than acceptable value but it is expected as there is high degree of correlation.

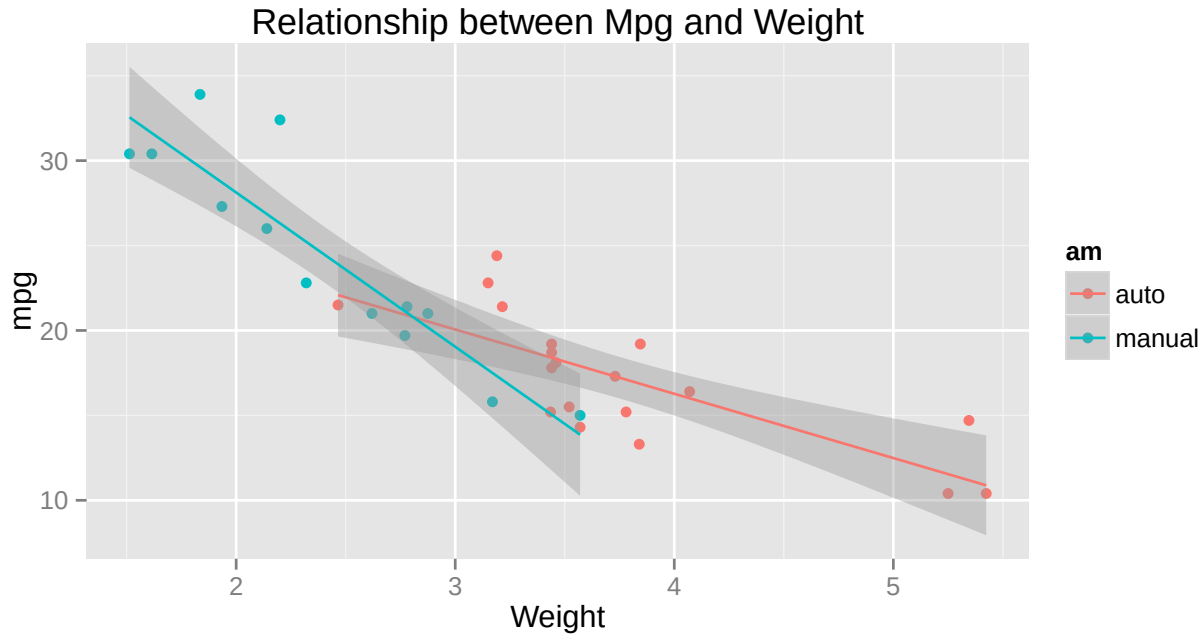
```
summary(fit1)$coefficients

##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.44019    9.31887   1.871 0.072149
## hp          -0.01765    0.01415  -1.247 0.223088
## wt          -3.23810    0.88990  -3.639 0.001141
## qsec         0.81060    0.43887   1.847 0.075731
## ammanual     2.92550    1.39715   2.094 0.045791

summary(fit2)$coefficients

##               Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 26.6248479    2.18294 12.19677 1.014e-12
## hp          -0.0591370    0.01294 -4.56838 9.019e-05
## ammanual     5.2176534    2.66509  1.95778 6.029e-02
## hp:ammanual  0.0004029    0.01646  0.02448 9.806e-01
```

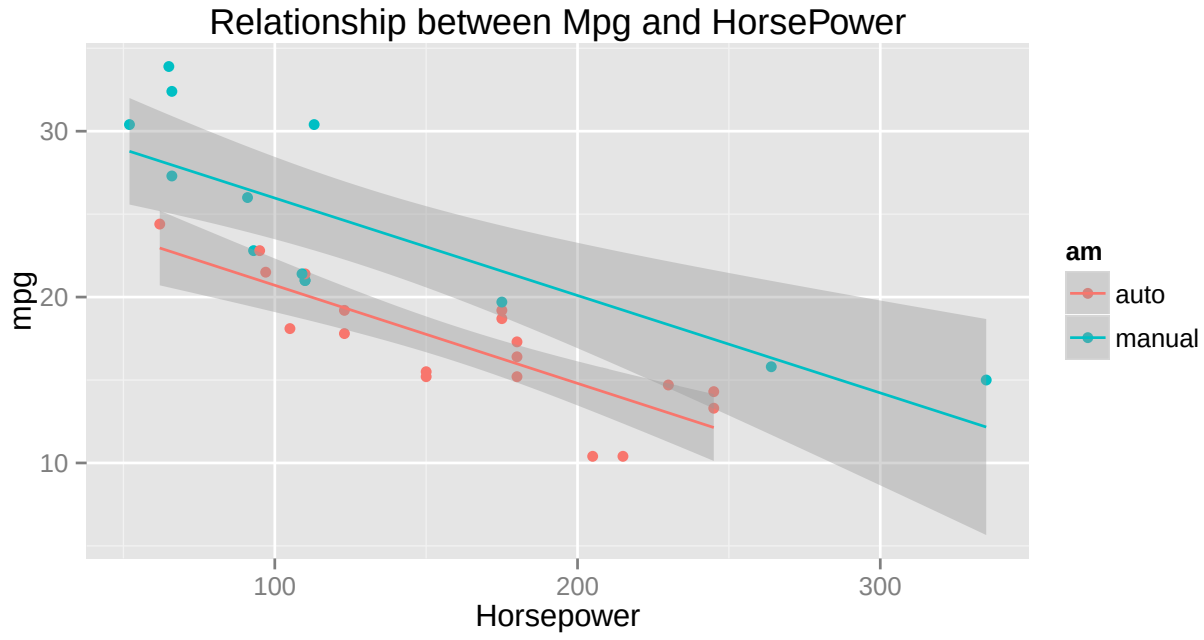
Both Models had a significant value of p for the 'am' variable. Hence both are credible.

Appendix

Part 1:
Exploratory Plots

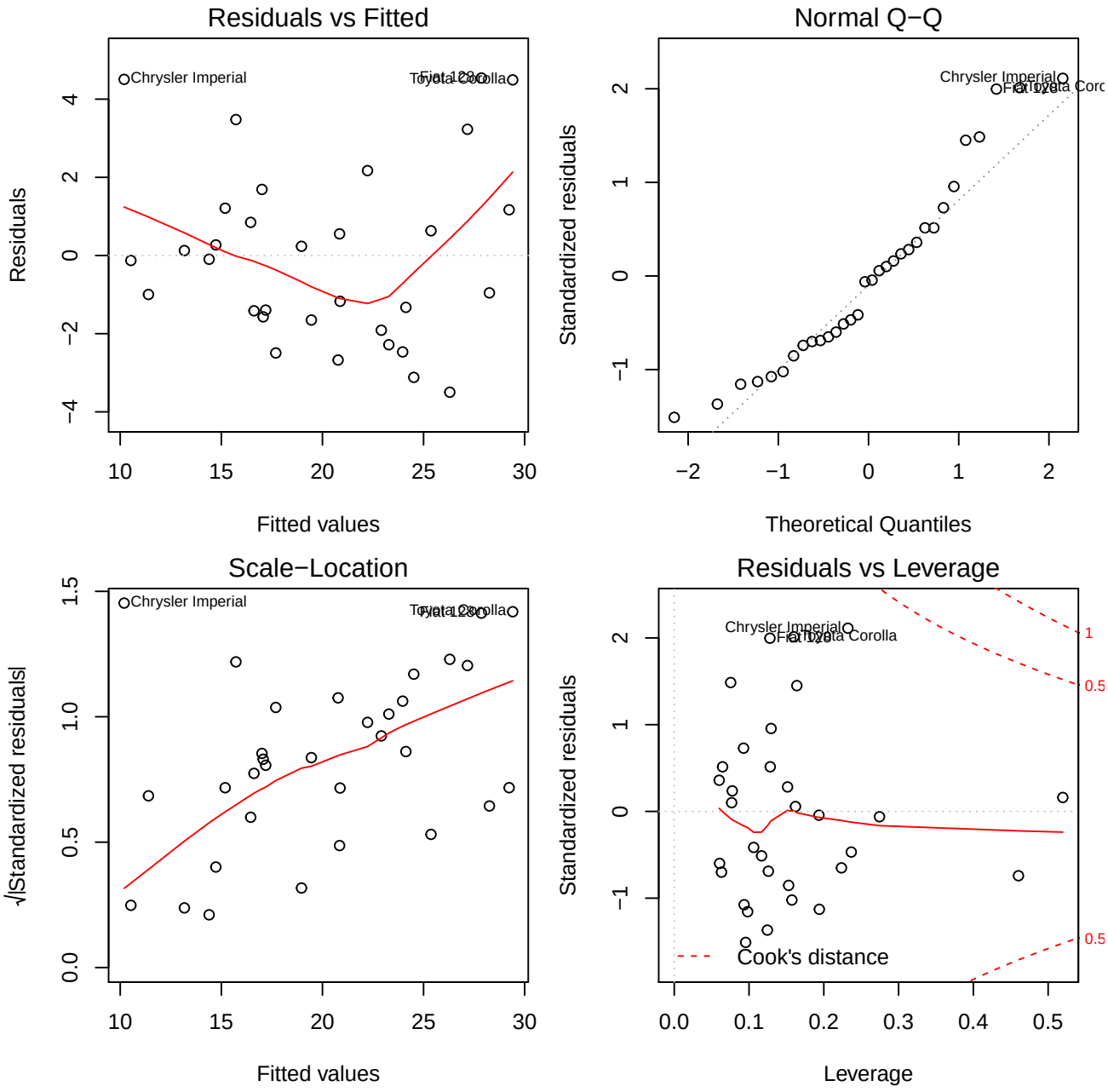
```
library(ggplot2)
ggplot(mtcars)+geom_point(aes(x=wt,y=mpg,colour=am))+geom_smooth(aes(x=wt,y=mpg,colour=am),
  method="lm")+labs(title="Relationship between Mpg and Weight",x="Weight")
ggplot(mtcars)+geom_point(aes(x=hp,y=mpg,colour=am))+geom_smooth(aes(x=hp,y=mpg,colour=am),
  method="lm")+labs(title="Relationship between Mpg and HorsePower",x="Horsepower")
```



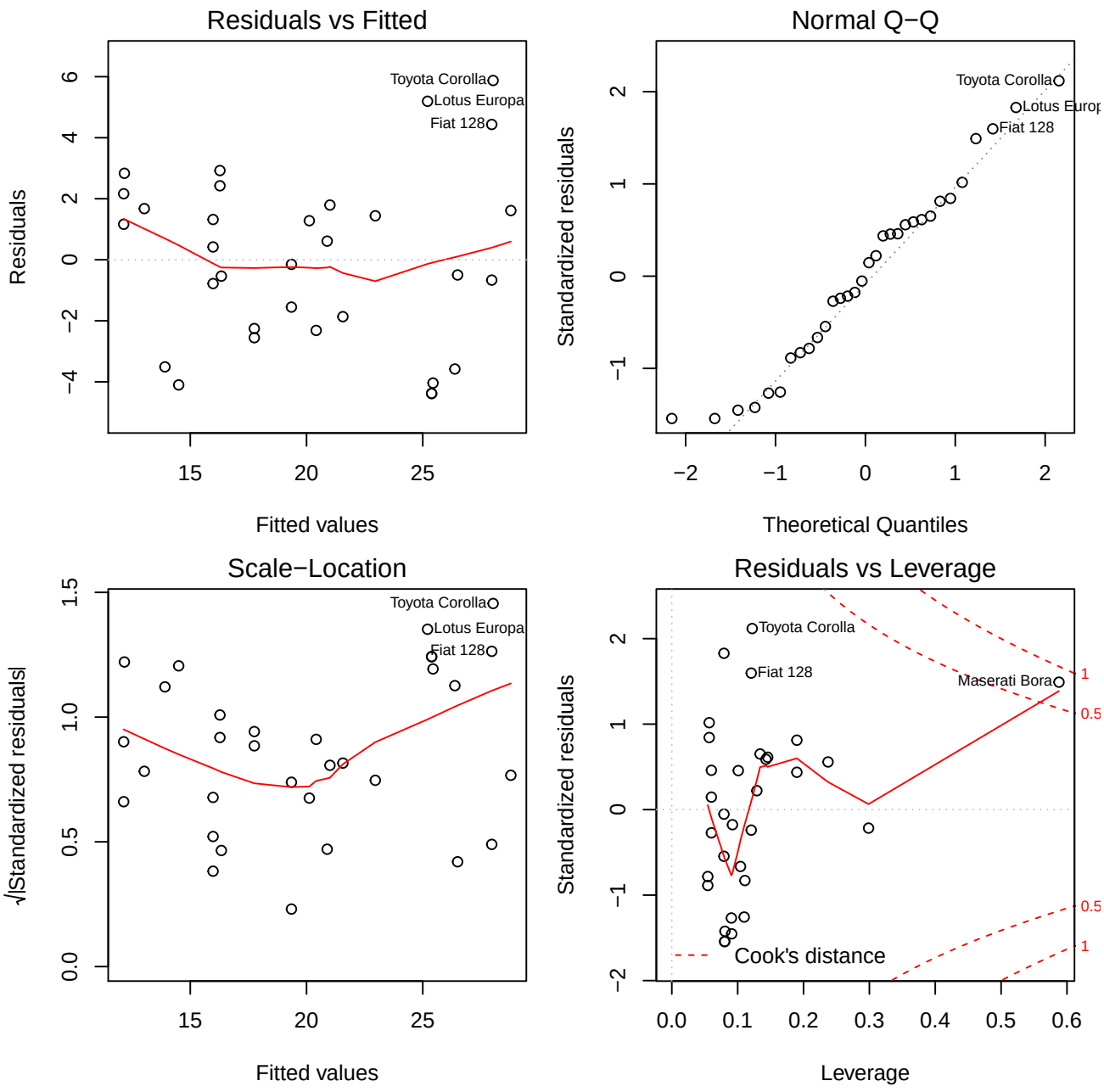


Part 2:
Residual Plots

```
par(mfrow=c(2,2),mar=c(4,4,2,1))
plot(fit1)
```




```
par(mfrow=c(2,2),mar=c(4,4,2,1))
plot(fit2)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Lotus Europa
Fiat 128
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Lotus Europ
Fiat 128
Scale−Location
√|Standardized residuals|
Fitted values
Toyota Corolla
Lotus Europa
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Fiat 128
Maserati Bora
Cook's distance
1
0.5
0.5
1