Regression Models Project

Jamshid Sodikov

September 24, 2015

Task

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?”

Question

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

Your report must be:

Written as a PDF printout of a compiled (using knitr) R markdown document. Brief. Roughly the equivalent of 2 pages or less for the main text. Supporting figures in an appendix can be included up to 5 total pages including the 2 for the main report. The appendix can only include figures. Include a first paragraph executive summary. Upload your PDF by clicking the Upload button below the text box.

Load Data

```
data(mtcars)
```

Exploratory analysis

```
summary(mtcars) #Summary statistics are omitted due to space limitation
```

```
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
```

```
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'))
```

Regression model

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

```
summary(best.model)$coeff
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## amManual     1.80921138 1.39630450  1.295714 2.064597e-01
```

Final model resulted in the following regressors such as the cyl6, cyl8, hp, wt, and amManual variables (overall p-value<0.001). The adjusted R-squared equals to 84% which is variance explained by the final model. Moreover, the results suggest 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).

```
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
```

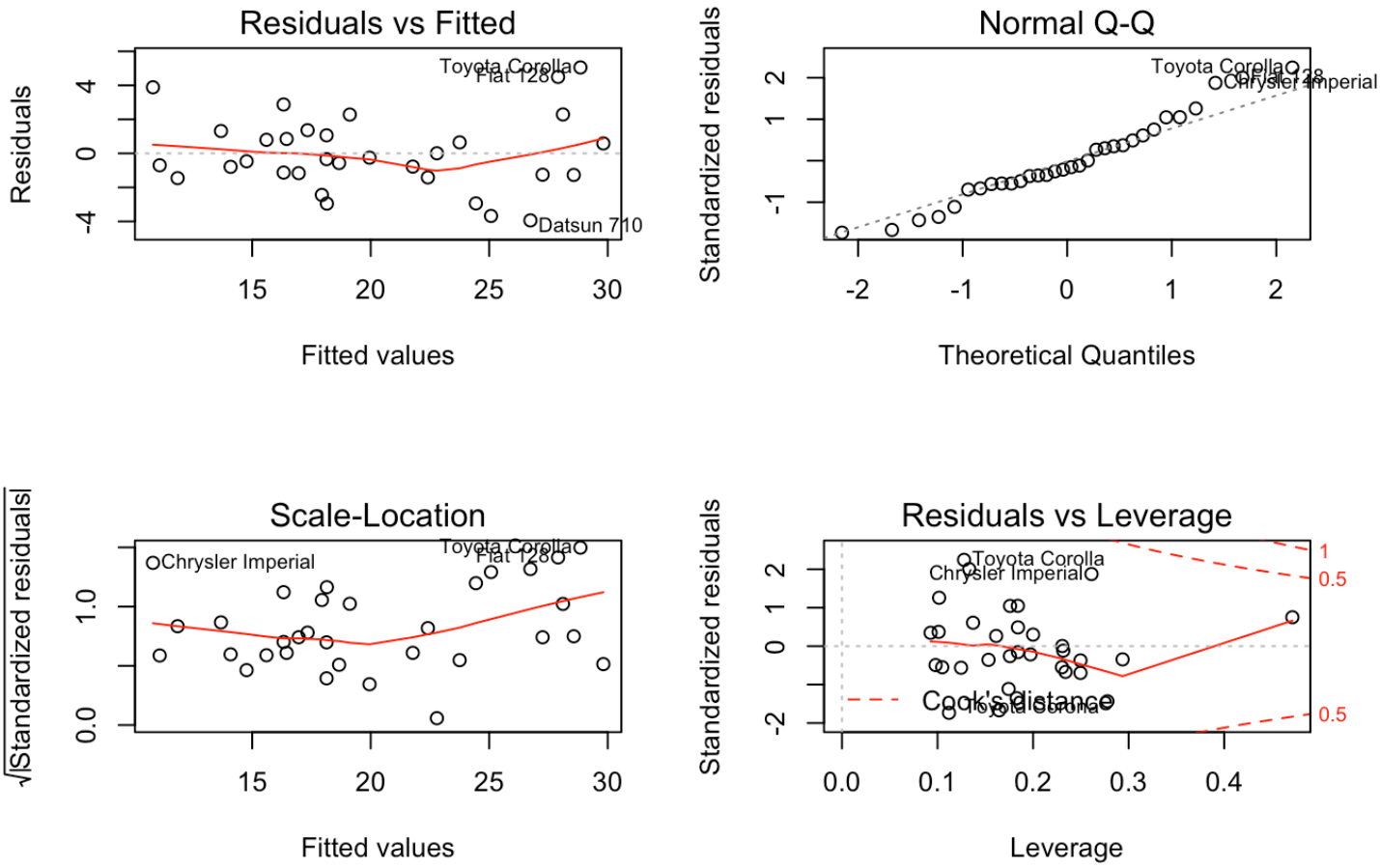
Conclusion

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.

Appendix

Residuals plot

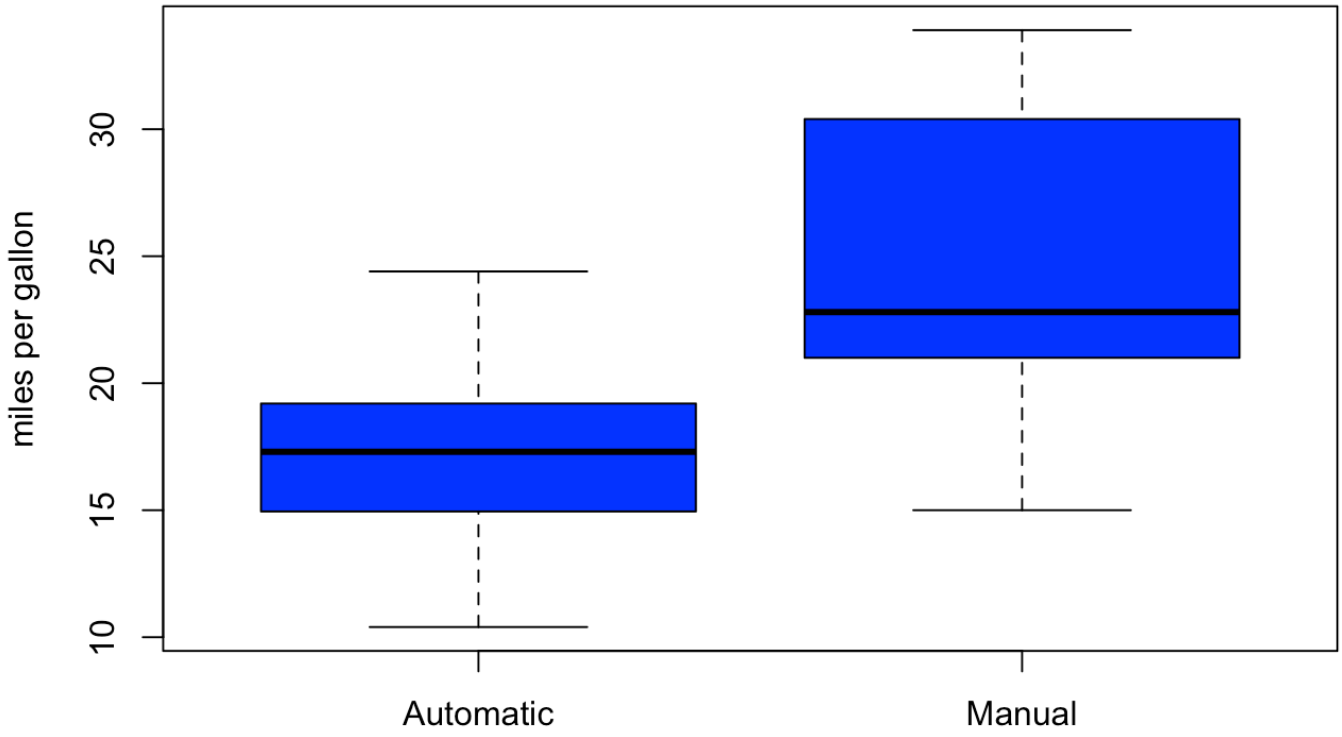
```
par(mfrow=c(2, 2))
plot(best.model)
```



Residual plots suggest that some transformation may be necessary to achieve linearity.

Box-plot

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



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).