

Regression Models - Course Project

Lorena Zuniga

Executive Summary

For this project, Motor Trend (a magazine about the automobile industry) is interested in exploring the relationship between a set of variables and miles per gallon (MPG). They need to answer the following questions: 1- Is an automatic or manual transmission better for MPG ? 2- Quantify the MPG difference between automatic and manual transmissions This project use the mtcars dataset

Exploratory Analysis

Load the data:

```
data("mtcars")
str(mtcars)
```

In the am column: 0 is automatic transmission, 1 = manual transmission. It is observed that the MPG mean between each type of transmission is different:

```
autom<- mtcars[mtcars$am==0,]
manual <-mtcars[mtcars$am==1,]
c(mean(autom$mpg),mean(manual$mpg))

## [1] 17.14737 24.39231
```

Inference

It is necessary to determine if the MPG mean difference is significant, then we can perform a t-test. Alpha value =0.05

```
t.test(autom$mpg,manual$mpg)
```

The p-value is 0.001374, so in this case the difference between the MPG mean of the cars with manual transmission and the automatic cars is significant

Regression Analysis

At first we cant test a simple linear regression model using am as predictor and mpg as outcome

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

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

The coefficient and intercept shows that manual transmission cars have 7.24 MPG and automatic transmission cars have 17.24 MPG. The R-squared = 0.338 , that shows the model only explains 33.8% of the variance.

Multivariable linear regression

In order to know which variables could be better to include in the multivariable linear regression model, we can get a correlation matrix, as follows:

```
cor(mtcars)[1,]
```

From the resultant matrix we can say that wt and hp could be tested as predictors

```
simpleModel <- lm(mpg ~ am,data=mtcars)
multiModel <- lm(mpg ~ am + hp + wt, data=mtcars)
```

Comparing both models with ANOVA:

```
anova(simpleModel, multiModel)
```

The p-value is 3.7e-09 and therefore the difference between both models is significant

```
summary(multiModel)$coef
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	34.00287512	2.642659337	12.866916	2.824030e-13
##	am	2.08371013	1.376420152	1.513862	1.412682e-01
##	hp	-0.03747873	0.009605422	-3.901830	5.464023e-04
##	wt	-2.87857541	0.904970538	-3.180850	3.574031e-03

Observing the coefficients, it is possible to say that, in average, cars with manual transmission have 2.083MPG more than cars with automatic transmission. For the multivariable model the R-squared = 0.84, it is possible to say that the multivariable model explain the 84% of the variance.

Residuals

It is verified that the mean of the residuals is near 0:

```
mean(multiModel$residuals)
```

```
## [1] 1.561251e-16
```

Also it is possible verify that each one of the MPG predictors (hp, wt and am) are uncorrelated with the residuals The residuals chart can be found in the Appendix

```
cov(multiModel$residuals,mtcars$hp)
```

```
## [1] -2.268436e-14
```

```
cov(multiModel$residuals,mtcars$wt)
```

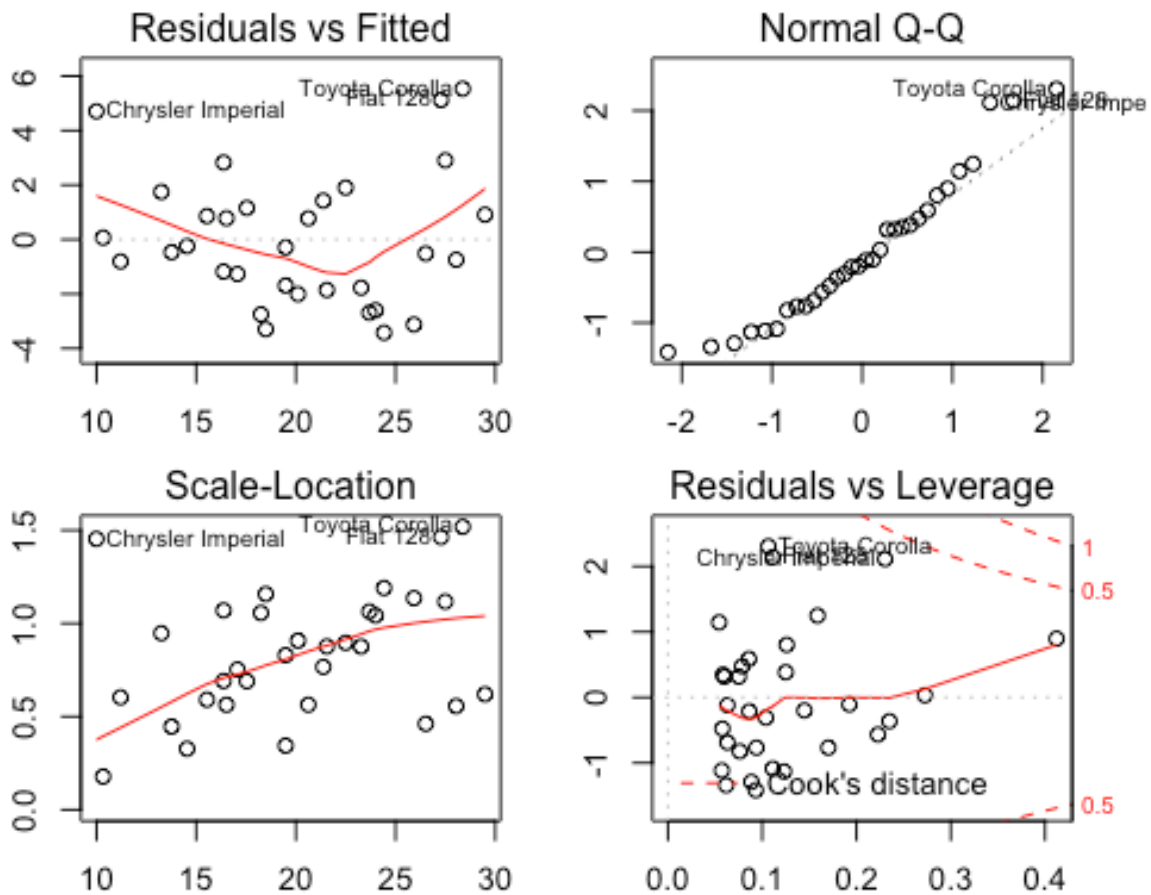
```
## [1] -3.79775e-16
```

```
cov(multiModel$residuals,mtcars$am)
```

```
## [1] 8.845691e-17
```

Appendix

```
par(mfrow=c(2,2))  
par(mar=rep(2,4))  
plot(multiModel)
```



The complete code for this project can be found here:

(<https://github.com/lorenzaz/regressionmodels/blob/master/Project/project.Rmd>)