Cars Prediction

Gherardo Lattanzi

2016-08-15

The Following is the initial dataset given for this Report. 50 Observations with 2 attributes each.

cars<- read.csv("~/Desktop/Ubiqum/Task2/cars.csv")  
knitr::kable(cars, align = "c")

|  |  |  |
| --- | --- | --- |
| name.of.car | speed.of.car | distance.of.car |
| Ford | 4 | 2 |
| Jeep | 4 | 4 |
| Honda | 7 | 10 |
| KIA | 7 | 10 |
| Toyota | 8 | 14 |
| BMW | 9 | 16 |
| Mercedes | 10 | 17 |
| GM | 10 | 18 |
| Hyundai | 10 | 20 |
| Infiniti | 11 | 20 |
| Land Rover | 11 | 22 |
| Lexus | 12 | 24 |
| Mazda | 12 | 26 |
| Mitsubishi | 12 | 26 |
| Nissan | 12 | 26 |
| GMC | 13 | 26 |
| Fiat | 13 | 28 |
| Chrysler | 13 | 28 |
| Dodge | 13 | 32 |
| Acura | 14 | 32 |
| Audi | 14 | 32 |
| Chevrolet | 14 | 34 |
| Buick | 14 | 34 |
| Ford | 15 | 34 |
| Jeep | 15 | 36 |
| Honda | 15 | 36 |
| KIA | 16 | 40 |
| Toyota | 16 | 40 |
| BMW | 17 | 42 |
| Mercedes | 17 | 46 |
| GM | 17 | 46 |
| Hyundai | 18 | 48 |
| Infiniti | 18 | 50 |
| Land Rover | 18 | 52 |
| Lexus | 18 | 54 |
| Mazda | 19 | 54 |
| Mitsubishi | 19 | 56 |
| Nissan | 19 | 56 |
| GMC | 20 | 60 |
| Fiat | 20 | 64 |
| Chrysler | 20 | 66 |
| Dodge | 20 | 68 |
| Acura | 20 | 70 |
| Audi | 22 | 76 |
| Chevrolet | 23 | 80 |
| Buick | 24 | 84 |
| Jeep | 24 | 85 |
| Honda | 24 | 92 |
| KIA | 24 | 93 |
| Dodge | 25 | 120 |

## DATA CLEANING

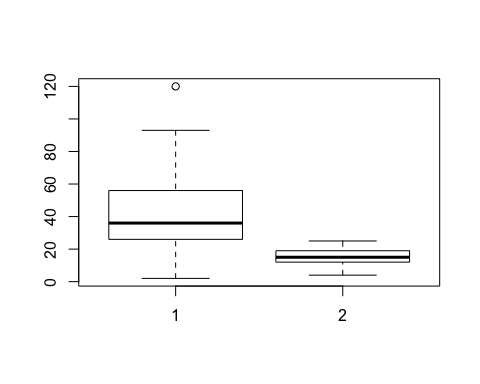
For the data cleaning work, the names of the colums are now in a technically correct form. Moreover, a new column called squared speed is created in order to spot the outliers.

### Create Squared Speed Column

names(cars) <-c("name.of.car","speed.of.car", "distance.of.car")  
cars$squared.speed <- cars$speed.of.car\*cars$speed.of.car

### Identify and Eliminate Outliers

boxplot(cars$distance.of.car, cars$speed.of.car)



outliers <- boxplot(cars$distance.of.car, plot=FALSE)$out  
show(outliers)

## [1] 120

###Outliers removed  
cars <- cars[-which(cars$distance.of.car %in% outliers),]

### Training and Testing TestSets

trainSize<-round(nrow(cars)\*0.7)   
print(trainSize)

## [1] 34

testSize<-nrow(cars)-trainSize  
print(testSize)

## [1] 15

training\_indices<-sample(seq\_len(nrow(cars)),size =trainSize)  
trainSet<-cars[training\_indices,]  
testSet<-cars[-training\_indices,]   
print(testSet)

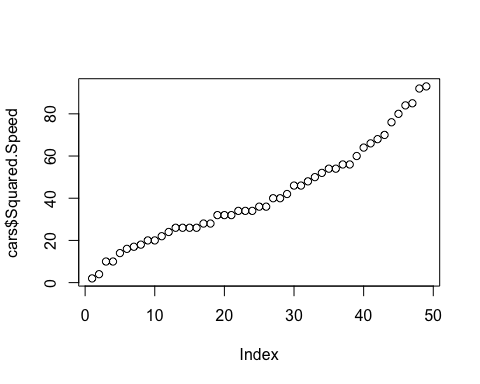
## name.of.car speed.of.car distance.of.car squared.speed  
## 4 KIA 7 10 49  
## 6 BMW 9 16 81  
## 7 Mercedes 10 17 100  
## 9 Hyundai 10 20 100  
## 10 Infiniti 11 20 121  
## 15 Nissan 12 26 144  
## 16 GMC 13 26 169  
## 24 Ford 15 34 225  
## 33 Infiniti 18 50 324  
## 34 Land Rover 18 52 324  
## 38 Nissan 19 56 361  
## 42 Dodge 20 68 400  
## 43 Acura 20 70 400  
## 47 Jeep 24 85 576  
## 48 Honda 24 92 576

### LINEAR REGRESSION

###Prediction

Prediction<- predict(MOD,testSet)  
  
knitr::kable(Prediction)

|  |  |
| --- | --- |
|  | x |
| 4 | 5.931049 |
| 6 | 14.489492 |
| 7 | 18.768713 |
| 9 | 18.768713 |
| 10 | 23.047935 |
| 15 | 27.327156 |
| 16 | 31.606378 |
| 24 | 40.164820 |
| 33 | 53.002485 |
| 34 | 53.002485 |
| 38 | 57.281706 |
| 42 | 61.560928 |
| 43 | 61.560928 |
| 47 | 78.677813 |
| 48 | 78.677813 |



### GRAPHIC Distance and Squared Speed