HW1

Doeun Kim

4/3/2018

Problem

I would like to prove whether the car funtionalities (horsepower, city MPG, highway MPG) directly affect the car price or not.

Gather and Pre-process the Data

To read the data properly and easily, I converted the given excel file '04cars data.xls' to the csv file '04cars.csv.' Moreover, in order to use valid data, I removed several records that have at least one asterisk from the funtionality features.

I extracted all the data of Retail price, Dealer price, Horsepower, City MPG, and Highway MPG from the csv file and put them in the corresponding array. The X axis always corresponds to the average price of the retail price and the dealer price because there are two types of price data and both types of data are precious, and the Y axis corresponds to the one feature among Horsepower, City MPG, and Highway MPG for each graph.

```
res = read.csv(file="04cars.csv", header=T)
retail_price = as.array(res[,10])
dealer_price = as.array(res[,11])
avg_price = (retail_price + dealer_price)/2
hp = as.array(res[,14])
city_mpg = as.array(res[,15])
hwy_mpg = as.array(res[,16])
```

Plot

Every graphs are plotted in Scatter Plot because what I want to see in the graphs is the relationship between the functionality and the price. Scatter Plot helps me to recognize the relationship through the regression line which is made based on the position of the data dots.

The first graph is telling you the relationship of the car price and the horsepower. By seeing the red non-linear regression line, you can recognize that as the horsepower becomes higher, the car price becomes higher.

```
#Plot the graph
x = as.numeric(avg_price)
y = as.numeric(hp)
plot(x, y, xlab = "Price($)", ylab = "HP(hp)", col = "blue", pch = 16, xlim =
```
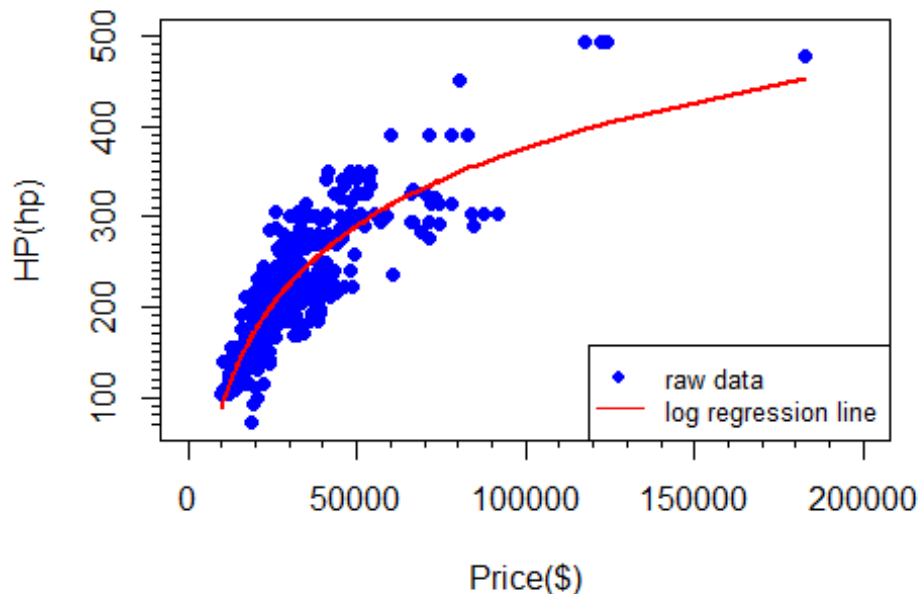
```
c(0,200000),ylim = c(70,500))
axis(side = 1, at = seq(0,200000, by = 10000),labels = FALSE, tcl = -0.2)
axis(side = 2, at = seq(70,500, by = 10),labels = FALSE, tcl = -0.2)

#Add the regression line on the graph
model = lm(y ~ log(x))
y = predict(model, newdata = list(x))
matlines(x, y, lwd = 2, col = "red")

#Add the legend on the graph
legend("bottomright",legend = c("raw data", "log regression line"), col =
c("blue", "red"), cex = 0.8, lty = c(NA, 1), pch = c(16, NA))
```


The second graph is telling you the relationship of the car price and the city MPG. By seeing the red non-linear regression line, you can recognize that as the city MPG becomes lower, the car price becomes higher. In other words, the cars that have high city MPG cost low.
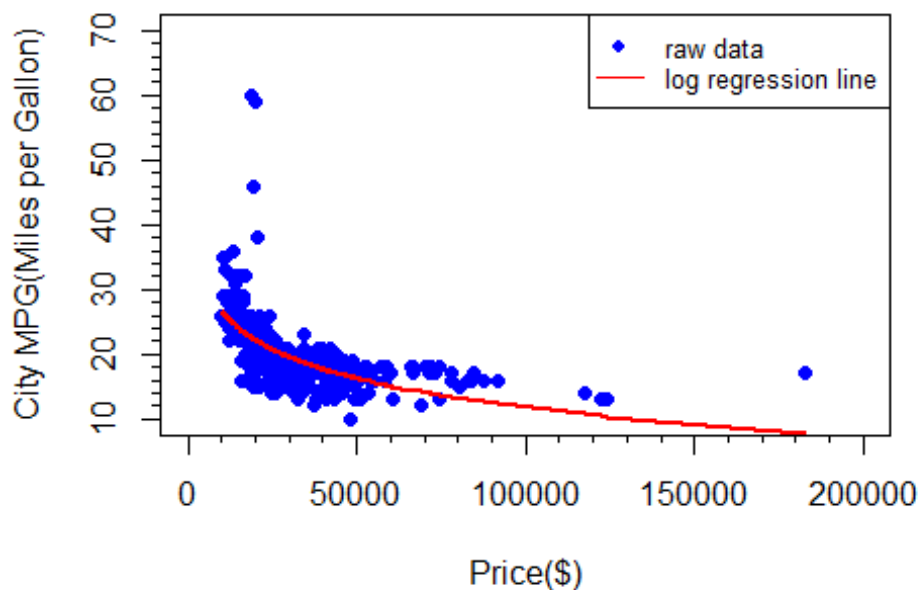
```
#Plot the graph
x = as.numeric(avg_price)
y = as.numeric(city_mpg)
plot(x, y, xlab = "Price($)", ylab = "City MPG(Miles per Gallon)", col =
"blue", pch = 16, xlim = c(0,200000) , ylim = c(10,70))
axis(side = 1, at = seq(0,200000, by = 10000),labels = FALSE, tcl = -0.2)
axis(side = 2, at = seq(10,70, by = 2),labels = FALSE, tcl = -0.2)
```

```
#Add the regression line on the graph
model = lm(y ~ log(x))
y = predict(model, newdata = list(x))
matlines(x, y, lwd = 2, col = "red")

#Add the legend on the graph
legend("topright",legend = c("raw data", "log regression line"), col =
c("blue", "red"), cex = 0.8, lty = c(NA, 1), pch = c(16, NA))
```


The third graph is telling you the relationship of the car price and the highway MPG. By seeing the red non-linear regression line, you can recognize that as the highway MPG becomes lower, the car price becomes higher. In other words, the cars that have high highway MPG cost low. This graph shows the same mechanism as the graph of City MPG vs Price.

```
#Plot the graph
x = as.numeric(avg_price)
y = as.numeric(hwy_mpg)
plot(x, y, xlab = "Price($)", ylab = "Highway MPG(Miles per Gallon)", col =
"blue", pch = 16, xlim = c(0,200000) , ylim = c(10,70))
axis(side = 1, at = seq(0,200000, by = 10000), labels = FALSE, tcl = -0.2)
axis(side = 2, at = seq(10,70, by = 2), labels = FALSE, tcl = -0.2)

#Add the regression line on the graph
model = lm(y ~ log(x))
```
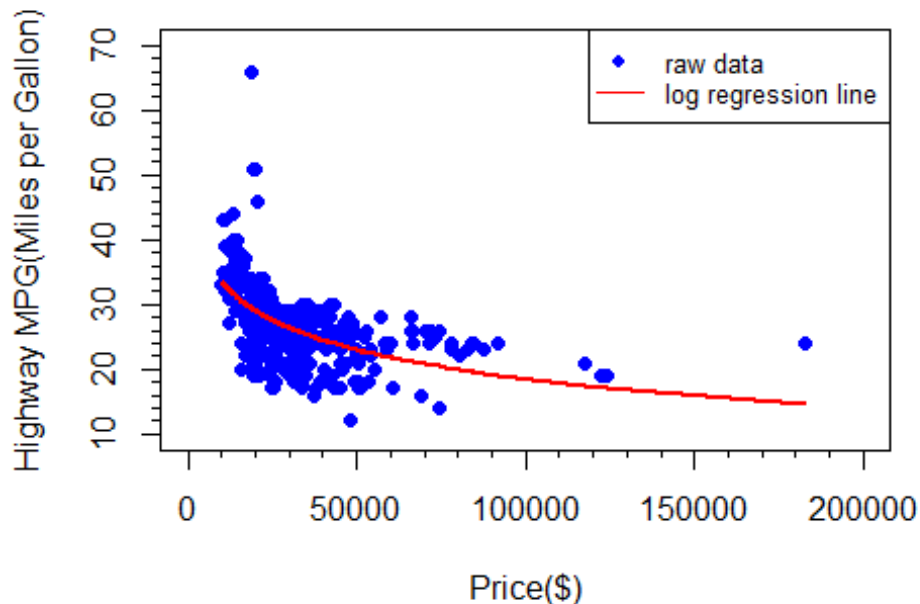
```
y = predict(model, newdata = list(x))
matlines(x, y, lwd = 2, col = "red")

#Add the legend on the graph
legend("topright",legend = c("raw data", "log regression line"), col =
c("blue", "red"), cex = 0.8, lty = c(NA, 1), pch = c(16, NA))
```


Conclusion

People typically think that a higher product price ensures a better product functionality. Therefore, before analyzing the given data, I thought that an expensive car has both high(good) city and highway MPGs and high horsepower. However, the graphs that I visualized proved that only the horsepower among three functionalities affects the car price.