Assignment 1

Rithin

2/29/2020

```
library(ISLR)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(tidyverse)

## -- Attaching packages ---------------------------------------------------
--------------------------------------------------- tidyverse 1.3.0 --

## v tibble  2.1.3     v purrr   0.3.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------
--------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

SafeBabies <- Carseats %>% select("Sales", "Price", "ShelveLoc")
head(SafeBabies)

##   Sales Price ShelveLoc
## 1  9.50   120       Bad
## 2 11.22    83      Good
## 3 10.06    80    Medium
## 4  7.40    97    Medium
## 5  4.15   128       Bad
## 6 10.81    72       Bad

good<-SafeBabies%>%filter(ShelveLoc=="Good")
head(good)
```

```
##   Sales Price ShelveLoc
## 1 11.22    83      Good
## 2 11.85   120      Good
## 3 11.96    94      Good
## 4 10.96    86      Good
## 5 11.17   118      Good
## 6  7.58   110      Good
```

```r
medium<-SafeBabies%>%filter(ShelveLoc=="Medium")
head(medium)
```

```
##   Sales Price ShelveLoc
## 1 10.06    80    Medium
## 2  7.40    97    Medium
## 3  6.63   108    Medium
## 4  6.54   124    Medium
## 5  4.69   124    Medium
## 6  3.98   136    Medium
```

```r
bad<-SafeBabies%>%filter(ShelveLoc=="Bad")
head(bad)
```

```
##   Sales Price ShelveLoc
## 1  9.50   120       Bad
## 2  4.15   128       Bad
## 3 10.81    72       Bad
## 4  9.01   100       Bad
## 5 10.14   113       Bad
## 6  2.99    97       Bad
```

model building

```r
## We are building linear model instead of building the genrelized linear
model because it makes no sense to build a genralized linear model on a
single variable

modelgood<-lm(Sales~Price,data=good)
summary(modelgood)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = good)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.721 -1.351 -0.098  1.483  4.353
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.968864   0.988008  18.187  < 2e-16 ***
## Price       -0.065785   0.008199  -8.023 5.85e-12 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.888 on 83 degrees of freedom
## Multiple R-squared:  0.4368, Adjusted R-squared:   0.43
## F-statistic: 64.37 on 1 and 83 DF,  p-value: 5.848e-12
```

```
modelbad<-lm(Sales~Price,data=bad)
summary(modelbad)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = bad)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4622 -1.0617 -0.2014  1.2050  4.6412
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.832984   0.990317  11.949  < 2e-16 ***
## Price       -0.055220   0.008486  -6.507  3.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.967 on 94 degrees of freedom
## Multiple R-squared:  0.3105, Adjusted R-squared:  0.3032
## F-statistic: 42.34 on 1 and 94 DF,  p-value: 3.702e-09
```

optimal cost for shelve location good

```
Productioncost<-55
paste("The optimal cost for shelve location good", ((-
modelgood$coefficients[[2]] *Productioncost) +
(modelgood$coefficients[[1]]))/(-2 * modelgood$coefficients[[2]]))
```

```
## [1] "The optimal cost for shelve location good 164.07312564386"
```

optimal cost for shelve location bad

```
paste("The optimal cost for shelve location bad", ((-
modelbad$coefficients[[2]] *Productioncost) +
(modelbad$coefficients[[1]]))/(-2 * modelbad$coefficients[[2]]))
```

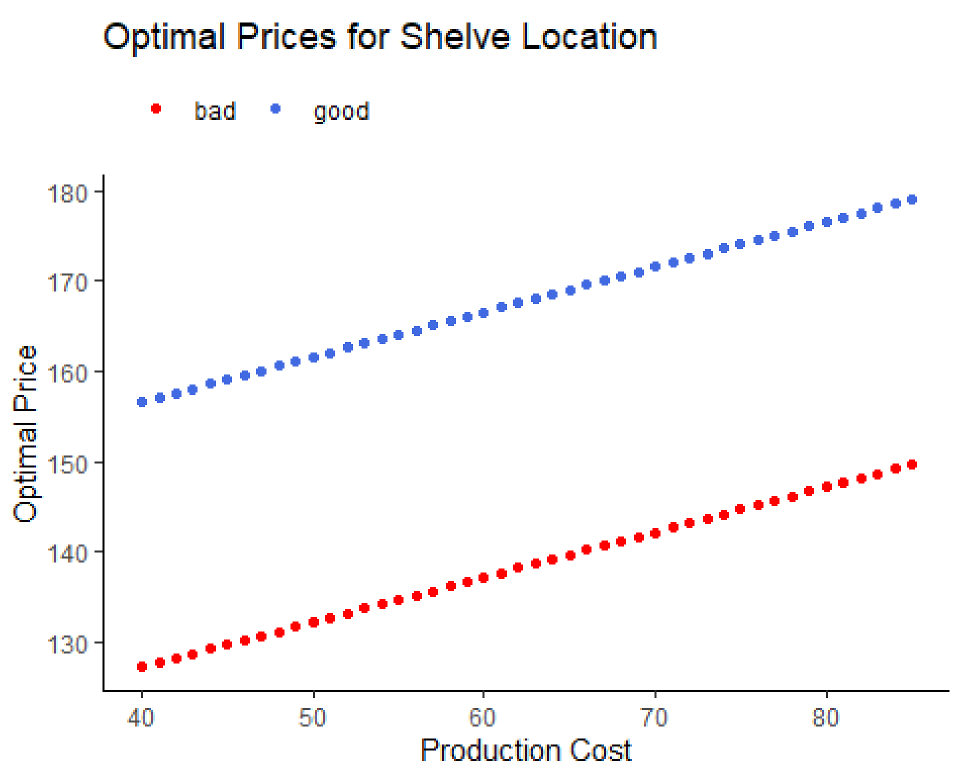
```
## [1] "The optimal cost for shelve location bad 134.643464696399"
```

plotting optimal cost for varying production costs

```
productioncost1<-data.frame(prodncost=seq(40,85,by=1),optimalprice=rep(0,46))
for (i in 1:nrow(productioncost1)) {
  productioncost1$optimalprice[i]<- ((-modelgood$coefficients[[2]]
*productioncost1$prodncost[i]) + (modelgood$coefficients[[1]]))/(-2 *
modelgood$coefficients[[2]])
}
```

```r
productioncost2 <- c(40:85)
rr <- 0
for (i in 1:length(productioncost2)) {
   rr[i] <- ((-modelbad$coefficients[[2]] *productioncost2[i]) +
(modelbad$coefficients[[1]]))/(-2 * modelbad$coefficients[[2]])
}
productioncost2<-data.frame(prodncost=seq(40,85,by=1),optimalprice=rr)
data<-cbind(productioncost1,productioncost2$optimalprice)
names(data)<-c("productioncost","good","bad")
data1<- data %>% gather("category","optimalprice",-productioncost)

ggplot(data1,aes(x=productioncost,y=optimalprice,color=category))+
  geom_point()+theme_classic()+ labs(x="Production Cost",y="Optimal
Price",color="Shelve Location") +
  ggtitle("Optimal Prices for Shelve Location")+
  theme(legend.position = "top",legend.justification = "left",legend.title =
element_blank()) + scale_color_manual(values=c("red1","royalblue"))
```




```r
## We can observe from the above plot that if the Shelve location is good the
optimal price for the product is greater than the products placed at shleve
location bad, That makes sense!!!
```