MIS-64037: Individual Assignment #1

Steven Spence

2/22/2019

SafeBabies is a large company who is producing car seats for babies and toddlers. They sell their products all over the US and abroad. The management team has hired you as a Business Analytics consultant to help them maximizing their profit.

Your primary task is to determine: 1. The optimal price for selling the car seats at those stores where the shelve location is good (i.e. the product is highly visible)? 2. The optimal price for selling the car seats at those stores where the shelve location is bad (i.e. the product is highly visible)?

You have been told that the cost of producing each car seat is $55.0

1. Plot the optimal price for selling the car seats at those stores where the shelve location is good and those where the shelve location is bad when varying the production costs from $40 to $85.

set.seed(1000)  
  
# Imports a subset of the Carseats dataframe from the ISLR library  
# Where:  
# "Sales" is the unit sales (in thousands) at each location  
# "Price" is the price that the company charges for each car seat  
# "ShelveLoc" indicate the shelving location of the car seat at each store (Factor w/ levels "Bad", "Good", and "Medium")  
  
require(ISLR)

## Loading required package: ISLR

# Create subset data frames for each type of shelving location  
  
SafeBabies <- Carseats[ , c("Sales", "Price", "ShelveLoc")]  
SafeBabies$Cost <- 55  
SafeBabies$Sales\_1K <- SafeBabies$Sales\*1000  
SafeBabies$Profit\_Per\_Unit <- SafeBabies$Price - SafeBabies$Cost  
SafeBabies$Total\_Profit <- SafeBabies$Sales \* SafeBabies$Profit\_Per\_Unit  
SafeBabies\_Good <- subset(SafeBabies, ShelveLoc == "Good")  
SafeBabies\_Bad <- subset(SafeBabies, ShelveLoc == "Bad")  
SafeBabies\_Medium <- subset(SafeBabies, ShelveLoc == "Medium")

Problem Statement:

The company would like to maximize their profit by optimizing their price based on shelf locations of their products around the world.

There we need to determine the price sensitivity of customers based on shelf location at various stores with historic data provided.

Once we know the relationship between price and sales, then we can find the optimum price, as well as plotting out optimum prices at various levels of manufacturing costs.

Reivew of Data Presented by Company:

# Summarize "SafeBabies" dataframe to review high level statistics for 400 stores  
  
summary(SafeBabies)

## Sales Price ShelveLoc Cost   
## Min. : 0.000 Min. : 24.0 Bad : 96 Min. :55   
## 1st Qu.: 5.390 1st Qu.:100.0 Good : 85 1st Qu.:55   
## Median : 7.490 Median :117.0 Medium:219 Median :55   
## Mean : 7.496 Mean :115.8 Mean :55   
## 3rd Qu.: 9.320 3rd Qu.:131.0 3rd Qu.:55   
## Max. :16.270 Max. :191.0 Max. :55   
## Sales\_1K Profit\_Per\_Unit Total\_Profit   
## Min. : 0 Min. :-31.0 Min. :-323.3   
## 1st Qu.: 5390 1st Qu.: 45.0 1st Qu.: 287.1   
## Median : 7490 Median : 62.0 Median : 399.9   
## Mean : 7496 Mean : 60.8 Mean : 426.1   
## 3rd Qu.: 9320 3rd Qu.: 76.0 3rd Qu.: 549.1   
## Max. :16270 Max. :136.0 Max. :1057.8

Visual Analysis of Data Presented:

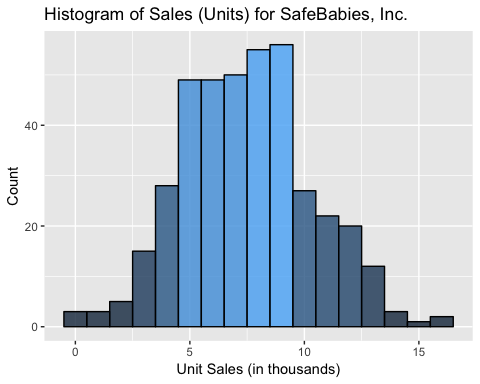
Company Sales:

# Create a histogram for the "Sales" variable  
  
require(ggplot2)

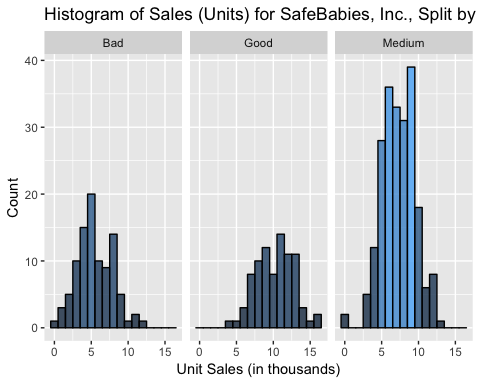
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.4.4

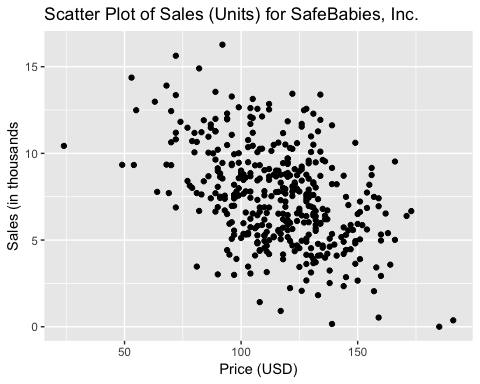
ggplot(data=SafeBabies, aes(SafeBabies$Sales)) +  
 geom\_histogram(col="black",  
 alpha=0.8,  
 aes(fill=..count..),  
 binwidth = 1) +  
 labs(title="Histogram of Sales (Units) for SafeBabies, Inc.") +  
 labs(x="Unit Sales (in thousands)", y="Count") +  
 guides(fill=FALSE)



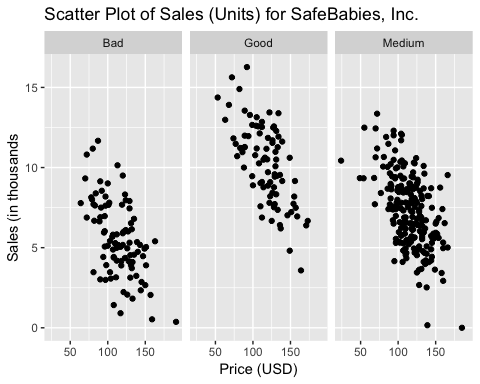
# Create a histogram for the "Sales" variable, split by the shelf location  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(SafeBabies$Sales)) +  
 geom\_histogram(col="black",  
 alpha=0.8,  
 aes(fill=..count..),  
 binwidth = 1) +  
 labs(title="Histogram of Sales (Units) for SafeBabies, Inc., Split by Shelf Location") +  
 labs(x="Unit Sales (in thousands)", y="Count") +  
 guides(fill=FALSE) +  
 facet\_wrap(vars(SafeBabies$ShelveLoc))



# Create a scatter plot for the "Sales" variable versus "Price" variable.  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(x = SafeBabies$Price)) +  
 geom\_point(aes(y= SafeBabies$Sales)) +  
 labs(title="Scatter Plot of Sales (Units) for SafeBabies, Inc.") +  
 labs(x="Price (USD)", y="Sales (in thousands")



# Create a scatter plot for the "Sales" variable versus "Price" variable, split by shelf location.  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(x = SafeBabies$Price)) +  
 geom\_point(aes(y= SafeBabies$Sales)) +  
 labs(title="Scatter Plot of Sales (Units) for SafeBabies, Inc.") +  
 labs(x="Price (USD)", y="Sales (in thousands") +  
 facet\_wrap(vars(SafeBabies$ShelveLoc))

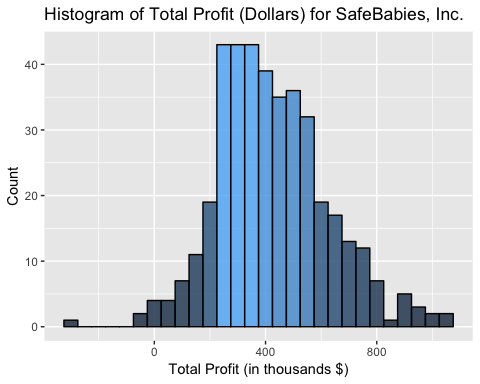


Review of the “Sales” charts we can state:

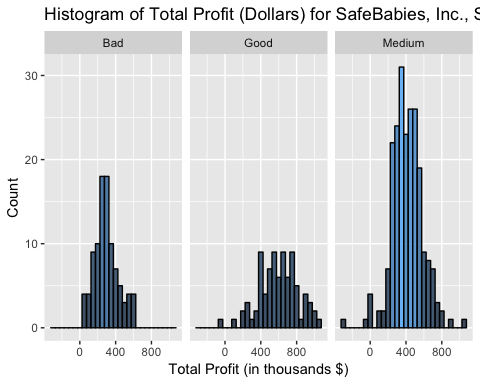
1. The data appears to be normally distributed
2. There appears to be a negative correlation between price and sales. Increasing price will result in less sales - as expected.

Total Profit:

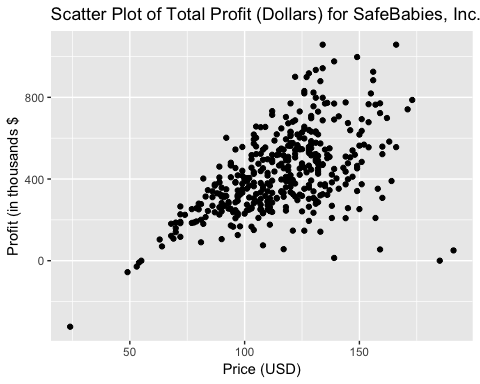
# Create a histogram for the "Total\_Profit" variable  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(SafeBabies$Total\_Profit)) +  
 geom\_histogram(col="black",  
 alpha=0.8,  
 aes(fill=..count..),  
 binwidth = 50) +  
 labs(title="Histogram of Total Profit (Dollars) for SafeBabies, Inc.") +  
 labs(x="Total Profit (in thousands $)", y="Count") +  
 guides(fill=FALSE)



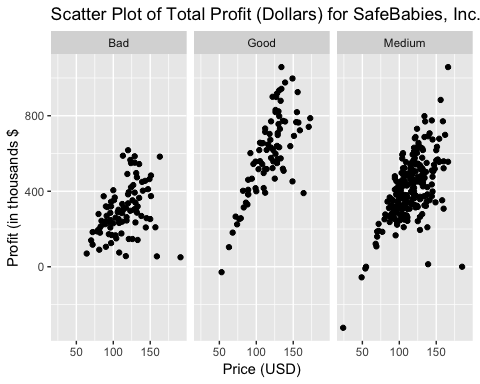
# Create a histogram for the "Total\_Profit" variable, split by the shelf location  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(SafeBabies$Total\_Profit)) +  
 geom\_histogram(col="black",  
 alpha=0.8,  
 aes(fill=..count..),  
 binwidth = 50) +  
 labs(title="Histogram of Total Profit (Dollars) for SafeBabies, Inc., Split by Shelf Location") +  
 labs(x="Total Profit (in thousands $)", y="Count") +  
 guides(fill=FALSE) +  
 facet\_wrap(vars(SafeBabies$ShelveLoc))



# Create a scatter plot for the "Total\_Profit" variable versus "Price" variable.  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(x = SafeBabies$Price)) +  
 geom\_point(aes(y= SafeBabies$Total\_Profit)) +  
 labs(title="Scatter Plot of Total Profit (Dollars) for SafeBabies, Inc.") +  
 labs(x="Price (USD)", y="Profit (in thousands $")



# Create a scatter plot for the "Total\_Profit" variable versus "Price" variable, split by shelf location.  
  
require(ggplot2)  
  
ggplot(data=SafeBabies, aes(x = SafeBabies$Price)) +  
 geom\_point(aes(y= SafeBabies$Total\_Profit)) +  
 labs(title="Scatter Plot of Total Profit (Dollars) for SafeBabies, Inc.") +  
 labs(x="Price (USD)", y="Profit (in thousands $") +  
 facet\_wrap(vars(SafeBabies$ShelveLoc))



Simple Linear Models - Good and Bad:

# Create simple linear model for good shelf location to get a baseline for the values.  
  
lm\_model\_good <- lm(Sales\_1K ~ Price, data = SafeBabies\_Good)  
summary(lm\_model\_good)

##   
## Call:  
## lm(formula = Sales\_1K ~ Price, data = SafeBabies\_Good)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3721 -1351 -98 1483 4353   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17968.864 988.008 18.187 < 2e-16 \*\*\*  
## Price -65.785 8.199 -8.023 5.85e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1888 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

# Create simple linear model for bad shelf location to get a baseline for the values.  
  
lm\_model\_bad <- lm(Sales\_1K ~ Price, data = SafeBabies\_Bad)  
summary(lm\_model\_bad)

##   
## Call:  
## lm(formula = Sales\_1K ~ Price, data = SafeBabies\_Bad)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4462.2 -1061.7 -201.4 1205.0 4641.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11832.984 990.317 11.949 < 2e-16 \*\*\*  
## Price -55.220 8.486 -6.507 3.7e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1967 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

Linear Regression for Good Shelf Location:

From this simple linear regression, we can see that the b0 (intercept) value was found to be 17968.864 and the b1 (price coefficient) was found to be -65.785.

Linear Regression for Bad Shelf Location:

From this simple linear regression, we can see that the b0 (intercept) value was found to be 11832.984 and the b1 (price coefficient) was found to be -55.220.

From the historic data, it is shown that there is a smaller drop in sales for every dollar increase at a bad shelf location compared to a good shelf location.

We can further improve this model by using a ridge and/or lasso method.

Generalized Linear Regression Models - Good and Bad:

# Cross validated linear regression model via lasso for the good shelf location  
  
require(glmnet)

## Loading required package: glmnet

## Warning: package 'glmnet' was built under R version 3.4.4

## Loading required package: Matrix

## Loading required package: foreach

## Loaded glmnet 2.0-16

glm\_model\_good <- cv.glmnet(x = cbind(0,SafeBabies\_Good$Price), y = SafeBabies\_Good$Sales\_1K, alpha = 1, nlambda = 100)  
  
coef(glm\_model\_good, s = "lambda.min")

## 3 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 17861.46755  
## V1 .   
## V2 -64.87373

# Cross validated linear regression model via lasso for the bad shelf location  
  
require(glmnet)  
  
glm\_model\_bad <- cv.glmnet(x = cbind(0,SafeBabies\_Bad$Price), y = SafeBabies\_Bad$Sales\_1K, alpha = 1, nlambda = 100)  
  
coef(glm\_model\_bad, s = "lambda.min")

## 3 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 11787.42010  
## V1 .   
## V2 -54.82154

The improved values for the model coefficient for good and bad are:

Good shelf location: -64.87373 Bad shelf location: -54.82154

Optimum Price for Good Shelving Location Stores:

We can use the previously gathered coefficient values to solve the following equation for optimum values:

Total Profit= (price – cost) \* (b1*price+b0) = b1*price^2+(b0 -cost*b1)*price-cost\*b0

When we plug the values in for good shelf location, the equation is:

b0\_good is 17968.864 b1\_good is -64.87373

Therefore,

-64.87373*price^2 + 21536.92*price - 988287.5

Taking the derivative will give the following equation:

-129.7475\*price + 21536.92

Setting the derivative equal to zero and solving for “price” will give us the optimal value in this situation. As a result, the optimum price for stores with a good shelf location will be:

Optimum Price for Good Shelf Location: $165.99

Optimum Price for Bad Shelving Location Stores:

We can use the previously gathered coefficient values to solve the following equation for optimum values:

Total Profit= (price – cost) \* (b1*price+b0) = b1*price^2+(b0 -cost*b1)*price-cost\*b0

When we plug the values in for good shelf location, the equation is:

b0\_bad is 11832.984 b1\_bad is -54.82154

Therefore,

-54.82154*price^2 + 14848.17*price - 650814.1

Taking the derivative will give the following equation:

-109.64308\*price + 14848.17

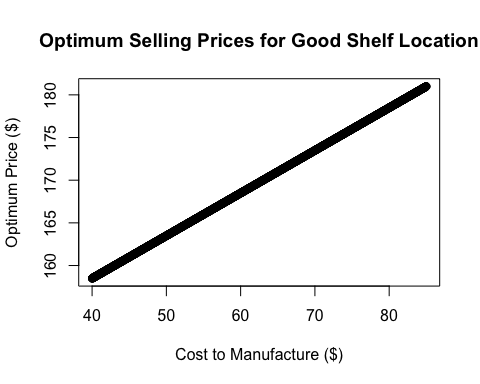
Setting the derivative equal to zero and solving for “price” will give us the optimal value in this situation. As a result, the optimum price for stores with a bad shelf location will be:

Optimum Price for Bad Shelf Location: $135.42

Plots of Optimum Price by Varying Costs:

Good Shelf Location:

# Use the derivative equation from before to substitute in all values between $40 and $85  
  
b0\_good <- 17968.864  
b1\_good <- -64.87373  
costs <- seq(from = 40, to = 85, by = 0.01)  
  
optimum\_prices\_good <- (-(b0\_good - costs \* b1\_good)) / (2\*b1\_good)  
  
plot(costs, optimum\_prices\_good, xlab = "Cost to Manufacture ($)", ylab = "Optimum Price ($)", main = "Optimum Selling Prices for Good Shelf Location")



Bad Shelf Location:

# Use the derivative equation from before to substitute in all values between $40 and $85  
  
b0\_bad <- 11832.984  
b1\_bad <- -54.82154  
costs <- seq(from = 40, to = 85, by = 0.01)  
  
optimum\_prices\_bad <- (-(b0\_bad - costs \* b1\_bad)) / (2\*b1\_bad)  
  
plot(costs, optimum\_prices\_bad, xlab = "Cost to Manufacture ($)", ylab = "Optimum Price ($)", main = "Optimum Selling Prices for Bad Shelf Location")

