Assignment3

Annie Beyer and Caleb Aguiar

1/29/2021

##Upload and Explore Data

load(file = "Detergent.RData")  
dt <- detergent\_DF  
library(skimr)  
library(psych)

## Warning: package 'psych' was built under R version 4.0.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.3

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

library(arm)

## Warning: package 'arm' was built under R version 4.0.3

## Loading required package: MASS

## Loading required package: Matrix

## Loading required package: lme4

##   
## arm (Version 1.11-2, built: 2020-7-27)

## Working directory is E:/MSBA\_UW/BUS\_AN\_514/R\_Code/analytics

##   
## Attaching package: 'arm'

## The following objects are masked from 'package:psych':  
##   
## logit, rescale, sim

library(broom)

## Warning: package 'broom' was built under R version 4.0.3

library(knitr)

## Warning: package 'knitr' was built under R version 4.0.3

skim(dt)

Data summary

|  |  |
| --- | --- |
| Name | dt |
| Number of rows | 14745 |
| Number of columns | 10 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 10 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| store | 0 | 1 | 80.99 | 35.81 | 2.00 | 53.00 | 86.00 | 111.00 | 139.00 | ▃▅▇▇▇ |
| week | 0 | 1 | 99.12 | 53.94 | 1.00 | 55.00 | 101.00 | 145.00 | 300.00 | ▆▇▇▁▁ |
| acv | 0 | 1 | 19159.97 | 4485.84 | 11141.43 | 15459.47 | 18966.02 | 22903.46 | 28003.89 | ▅▇▅▆▃ |
| promoflag | 0 | 1 | 0.82 | 0.39 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | ▂▁▁▁▇ |
| q\_tide128 | 0 | 1 | 81.22 | 134.14 | 1.00 | 30.00 | 48.00 | 78.00 | 2224.00 | ▇▁▁▁▁ |
| p\_tide128 | 0 | 1 | 8.36 | 0.76 | 4.88 | 7.95 | 8.48 | 8.87 | 10.51 | ▁▁▅▇▁ |
| q\_tide64 | 0 | 1 | 73.28 | 134.30 | 1.00 | 30.00 | 46.00 | 69.00 | 6906.00 | ▇▁▁▁▁ |
| p\_tide64 | 0 | 1 | 4.38 | 0.40 | 1.99 | 4.19 | 4.42 | 4.62 | 5.57 | ▁▁▂▇▁ |
| q\_wisk64 | 0 | 1 | 51.31 | 83.68 | 1.00 | 18.00 | 27.00 | 46.00 | 2309.00 | ▇▁▁▁▁ |
| p\_wisk64 | 0 | 1 | 4.07 | 0.49 | 2.03 | 3.79 | 4.19 | 4.40 | 7.72 | ▁▇▃▁▁ |

head(dt)

## store week acv promoflag q\_tide128 p\_tide128 q\_tide64 p\_tide64 q\_wisk64  
## 1 2 1 13828.87 1 34 8.701765 26 5.111538 71  
## 2 2 2 13828.87 0 46 8.670435 42 5.020476 16  
## 3 2 3 13828.87 0 43 8.720233 48 5.049375 35  
## 4 2 4 13828.87 0 165 8.738484 33 5.056667 26  
## 5 2 5 13828.87 1 77 6.990000 35 5.116857 33  
## 6 2 6 13828.87 1 45 8.656667 35 5.041429 37  
## p\_wisk64  
## 1 3.29  
## 2 4.19  
## 3 4.19  
## 4 4.19  
## 5 4.19  
## 6 4.23

## Question 1

# A

#Make Revenue column for each product  
  
dt$T128Revenue <- dt$q\_tide128 \* dt$p\_tide128  
dt$T64Revenue <- dt$q\_tide64 \* dt$p\_tide64  
dt$WiskRevenue <- dt$q\_wisk64 \* dt$p\_wisk64  
  
TotalRevenue <- sum(dt$T128Revenue) + sum(dt$T64Revenue) + sum(dt$WiskRevenue)  
  
T128Rev <- (sum(dt$T128Revenue) / TotalRevenue) \* 100  
T64Rev <- (sum(dt$T64Revenue) / TotalRevenue) \* 100  
WiskRev <- (sum(dt$WiskRevenue) / TotalRevenue) \* 100  
  
TotalRevenue1 <- format(round(TotalRevenue, 2), nsmall = 2)  
T128Rev1 <- format(round(T128Rev, 2), nsmall = 2)  
T64Rev1 <- format(round(T64Rev, 2), nsmall = 2)  
WiskRev1 <- format(round(WiskRev, 2), nsmall = 2)  
  
Mean128 <- mean(dt$p\_tide128)  
Mean64 <- mean(dt$p\_tide64)  
MeanW <- mean(dt$p\_wisk64)  
  
SD128 <- sd(dt$p\_tide128)  
SD64 <- sd(dt$p\_tide64)  
SDW <- sd(dt$p\_wisk64)  
  
Med128 <- median(dt$p\_tide128)  
Medn64 <- median(dt$p\_tide64)  
MednW <- median(dt$p\_wisk64)  
  
library(data.table)  
Marketshare = c(T128Rev1, T64Rev1, WiskRev1)  
MeanPrice = c(Mean128, Mean64, MeanW)  
MedianPrice = c(Med128, Medn64, MednW)  
Std.Dev = c(SD128, SD64, SDW)  
Product <- c("Tide 128 oz", "Tide 64 oz", "Wisk 64 oz")  
  
MarketShare <- data.frame(Product, Marketshare, MeanPrice, MedianPrice, Std.Dev)  
  
describe(dt)

## vars n mean sd median trimmed mad min  
## store 1 14745 80.99 35.81 86.00 83.38 40.03 2.00  
## week 2 14745 99.12 53.94 101.00 99.75 66.72 1.00  
## acv 3 14745 19159.97 4485.84 18966.02 19038.94 5512.81 11141.43  
## promoflag 4 14745 0.82 0.39 1.00 0.90 0.00 0.00  
## q\_tide128 5 14745 81.22 134.14 48.00 55.67 31.13 1.00  
## p\_tide128 6 14745 8.36 0.76 8.48 8.40 0.67 4.88  
## q\_tide64 7 14745 73.28 134.30 46.00 49.38 28.17 1.00  
## p\_tide64 8 14745 4.38 0.40 4.42 4.40 0.34 1.99  
## q\_wisk64 9 14745 51.31 83.68 27.00 32.47 17.79 1.00  
## p\_wisk64 10 14745 4.07 0.49 4.19 4.10 0.54 2.03  
## T128Revenue 11 14745 639.13 939.32 405.63 462.40 257.33 6.99  
## T64Revenue 12 14745 296.05 435.48 204.17 217.62 118.96 4.19  
## WiskRevenue 13 14745 188.92 263.64 112.07 130.67 68.21 3.79  
## max range skew kurtosis se  
## store 139.00 137.00 -0.46 -0.70 0.29  
## week 300.00 299.00 -0.02 -0.91 0.44  
## acv 28003.89 16862.46 0.16 -1.08 36.94  
## promoflag 1.00 1.00 -1.65 0.73 0.00  
## q\_tide128 2224.00 2223.00 7.43 77.56 1.10  
## p\_tide128 10.51 5.63 -0.47 0.18 0.01  
## q\_tide64 6906.00 6905.00 13.40 484.37 1.11  
## p\_tide64 5.57 3.58 -0.68 1.32 0.00  
## q\_wisk64 2309.00 2308.00 5.97 68.47 0.69  
## p\_wisk64 7.72 5.68 -0.09 1.51 0.00  
## T128Revenue 15579.56 15572.57 7.00 70.23 7.74  
## T64Revenue 13760.94 13756.75 7.22 99.36 3.59  
## WiskRevenue 7010.00 7006.21 5.60 62.32 2.17

MarketShare

## Product Marketshare MeanPrice MedianPrice Std.Dev  
## 1 Tide 128 oz 56.86 8.363219 8.476207 0.7600493  
## 2 Tide 64 oz 26.34 4.375451 4.419412 0.4046076  
## 3 Wisk 64 oz 16.81 4.071963 4.190000 0.4904530

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Product | Marketshare | Mean Price | Median Price | Std. Dev |
| Tide 128 oz | 56.9% | 8.36 | 8.48 | .76 |
| Tide 64 oz | 26.3% | 4.36 | 4.42 | .40 |
| Wisk 64 oz | 16.8% | 4.07 | 4.19 | .49 |

# B

Price Gap

dt$PriceGap1 <- dt$p\_tide128 - dt$p\_tide64  
dt$PriceGap2 <- dt$p\_tide64 - dt$p\_wisk64  
  
MeanPG1 <- mean(dt$PriceGap1)  
MeanPG2 <- mean(dt$PriceGap2)  
  
SDPG1 <- sd(dt$PriceGap1)  
SDPG2 <- sd(dt$PriceGap2)  
  
MedPG1 <- median(dt$PriceGap1)  
MedPG2 <- median(dt$PriceGap2)  
  
MeanPriceGap = c(MeanPG1, MeanPG2)  
MedianPriceGap = c(MedPG1, MedPG2)  
Std.DevPriceGap = c(SDPG1, SDPG2)   
PriceGap <- c("Tide128- Tide64", "Tide64 - Wisk64")   
  
PriceGap <- data.frame(PriceGap, MeanPriceGap, MedianPriceGap, Std.DevPriceGap)  
  
PriceGap

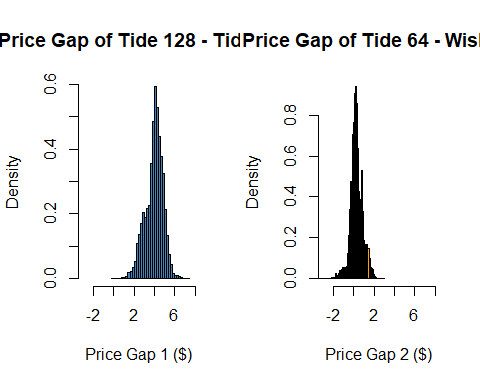
## PriceGap MeanPriceGap MedianPriceGap Std.DevPriceGap  
## 1 Tide128- Tide64 3.9877679 4.094000 0.8705744  
## 2 Tide64 - Wisk64 0.3034887 0.260714 0.5859641

|  |  |  |  |
| --- | --- | --- | --- |
| PriceGap | MeanPriceGap | MedianPrGap | Std. Dev |
| Tide128-Tide64 | 3.9877679 | 4.094000 | 0.8705 |
| Tide64 -Wisk64 | 0.3034887 | 0.260714 | | 0.5859 |

# C

Histograms

par(mfrow=c(1,2))#Set the plotting area into a 1\*2 array  
  
#Set axis to the same for better comparison  
hist(dt$PriceGap1, freq = F, col = "steelblue", breaks = 50, xlab = "Price Gap 1 ($)", main = "Price Gap of Tide 128 - Tide 64",  
 xlim = c(-3,8))  
hist(dt$PriceGap2, freq = F, col = "orange", breaks = 50, xlab = "Price Gap 2 ($)", main = "Price Gap of Tide 64 - Wisk 64",xlim = c(-3,8))



par(mfrow=c(1,1))

# D

Discussion: What do you learn from the price gap histograms and summary statistics for your analysis above?

Examining the revenue market share, Tide 128oz is about double the price, and double the amount of revenue of Tide 64oz.

This would suggest that the number of purchases made of each product is the same.

Tide 64 is more expensive than Wisk 64 on average.

The variability between Tide 128 and Tide 64 is less than the variability between Tide 64 and Wisk 64.

Is there enough variation in the price gaps across stores and weeks to estimate the cross price elasticities between the two Tide pack sizes and Wisk 64?

Yes, there is enough variation between each scenario.

## Question 2

# A

Sales velocity for Tide 128 and Tide 64

dt$svelocity128 <- (dt$q\_tide128/dt$acv)  
dt$svelocity64 <- (dt$q\_tide64/dt$acv)  
  
SalesVelocity128 <- sum(dt$svelocity128)  
SalesVelocity64 <- sum(dt$svelocity64)  
  
SalesVelocity128

## [1] 63.42925

SalesVelocity64

## [1] 56.98548

# B

Discussion: What is the purpose of dividing unit sales by ACV to construct the dependent variable?

The ACV is a measure of how large the store is. By dividing by ACV, this allows us to normalize the data so we can compare the prices of two products equally. We would expect a larger store to be more sensitive to price changes, as customers are more interested in price at a large store and have more options to substitute with, while customers at a small store are more likely to prioritize the convenience of the location and have less of a selection. Additionally, not all stores may sell each product every single week (whether that is due to operational supply difficulties, a lack of data recording, etc…)

# C

Log linear demand model

lm128 = lm(log(svelocity128) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64), data = dt)  
lm64 = lm(log(svelocity64) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64), data = dt)  
  
summary(lm128)

##   
## Call:  
## lm(formula = log(svelocity128) ~ log(p\_tide128) + log(p\_tide64) +   
## log(p\_wisk64), data = dt)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.0035 -0.4324 -0.0089 0.4180 2.9586   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.22420 0.16717 19.286 < 2e-16 \*\*\*  
## log(p\_tide128) -4.59708 0.06359 -72.288 < 2e-16 \*\*\*  
## log(p\_tide64) 0.28673 0.06142 4.668 3.07e-06 \*\*\*  
## log(p\_wisk64) 0.15141 0.04843 3.126 0.00177 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7166 on 14741 degrees of freedom  
## Multiple R-squared: 0.2662, Adjusted R-squared: 0.266   
## F-statistic: 1782 on 3 and 14741 DF, p-value: < 2.2e-16

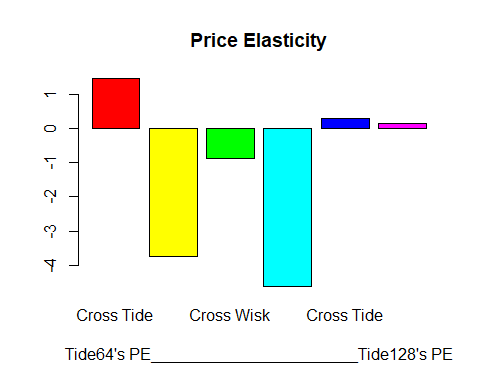
summary(lm64)

##   
## Call:  
## lm(formula = log(svelocity64) ~ log(p\_tide128) + log(p\_tide64) +   
## log(p\_wisk64), data = dt)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.9671 -0.3210 0.0805 0.4386 2.9940   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.35400 0.18165 -12.96 <2e-16 \*\*\*  
## log(p\_tide128) 1.44790 0.06910 20.95 <2e-16 \*\*\*  
## log(p\_tide64) -3.74860 0.06674 -56.17 <2e-16 \*\*\*  
## log(p\_wisk64) -0.87554 0.05263 -16.64 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7787 on 14741 degrees of freedom  
## Multiple R-squared: 0.2218, Adjusted R-squared: 0.2217   
## F-statistic: 1401 on 3 and 14741 DF, p-value: < 2.2e-16

lm128\_coef<- summary(lm128)$coefficients  
lm64\_coef<-summary(lm64)$coefficients#Lets graph these coefficients so we can see all these at a glance  
  
pe\_graph<-c(lm64\_coef[2:4,1],lm128\_coef[2:4,1])  
pe\_graph

## log(p\_tide128) log(p\_tide64) log(p\_wisk64) log(p\_tide128) log(p\_tide64)   
## 1.4478967 -3.7485955 -0.8755378 -4.5970780 0.2867306   
## log(p\_wisk64)   
## 0.1514120

barplot(pe\_graph, main = "Price Elasticity", xlab = "Tide64's PE\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Tide128's PE", col = rainbow(6), names.arg = c('Cross Tide','Own', 'Cross Wisk','Own', 'Cross Tide', 'Cross Wisk' ))



# D

Discussion: Discuss whether the demand estimates (own and cross price elasticities) make sense. Are the magnitudes and signs of the estimated parameters as you would expect?

OwnPE\_128 = -4.59 Elastic, negative OwnPE\_64 = -3.74 Elastic, negative

CrossPE\_128 = 0.31 Inelastic, positive. This is similar to addictive substances CrossPE\_64 = 1.28 Elastic Cross\_PE\_128 & Wisk = 0.15 Inelastic, positive Cross\_PE\_64 & Wisk = -0.87 Inelastic, negative

This regression shows that Tide128 and Tide64 are highly elastic, and there cross price elasticities are also elastic, but not as much. This suggests that as Tides own price increases, it’s demand goes down by a lot and Tide 64’s prices increases a little as well as the price for Wisk, but not as much. However, when Tide 64’s demand increases, its own price decreases, and Tide128 prices increase but only slightly. However, Wisk’s price decreases.

This does not make a lot of sense, because their price gap is sometimes at zero, and if that is the case customers are going to be falling over themselves to buy the Tide128 and not bother with Tide64. We need to examine if there is something going on here, is there more affluent locations where Tide64 is selling the same as Tide128 at a lower income area, or is there a seasonal element where people buy more Tide when they wash more clothes from sweating.

## Question 3

# A

Time Trend

dt$week<-as.integer(dt$week)  
lm128\_B <- lm(log(svelocity128) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64) + week, data = dt)  
lm64\_B <- lm(log(svelocity64) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64) + week, data = dt)  
  
lm128\_Ba<-tidy(lm128\_B)   
lm64\_Ba<-tidy(lm64\_B)  
  
kable(lm128\_Ba[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | 3.1776 | 0.1645 | 19.3211 | 0 |
| log(p\_tide128) | -4.7717 | 0.0630 | -75.6810 | 0 |
| log(p\_tide64) | 0.2935 | 0.0604 | 4.8566 | 0 |
| log(p\_wisk64) | 0.6295 | 0.0523 | 12.0408 | 0 |

kable(lm64\_Ba[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -2.4730 | 0.1646 | -15.0227 | 0 |
| log(p\_tide128) | 1.0020 | 0.0631 | 15.8777 | 0 |
| log(p\_tide64) | -3.7314 | 0.0605 | -61.6988 | 0 |
| log(p\_wisk64) | 0.3455 | 0.0523 | 6.6021 | 0 |

summary(lm128\_B)

##   
## Call:  
## lm(formula = log(svelocity128) ~ log(p\_tide128) + log(p\_tide64) +   
## log(p\_wisk64) + week, data = dt)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.9801 -0.4196 0.0012 0.4007 3.0205   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.1776253 0.1644641 19.321 < 2e-16 \*\*\*  
## log(p\_tide128) -4.7716627 0.0630496 -75.681 < 2e-16 \*\*\*  
## log(p\_tide64) 0.2934515 0.0604229 4.857 1.21e-06 \*\*\*  
## log(p\_wisk64) 0.6294800 0.0522789 12.041 < 2e-16 \*\*\*  
## week -0.0026325 0.0001185 -22.214 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7049 on 14740 degrees of freedom  
## Multiple R-squared: 0.29, Adjusted R-squared: 0.2898   
## F-statistic: 1505 on 4 and 14740 DF, p-value: < 2.2e-16

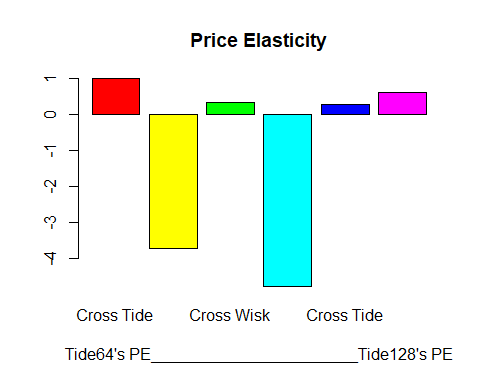
summary(lm64\_B)

##   
## Call:  
## lm(formula = log(svelocity64) ~ log(p\_tide128) + log(p\_tide64) +   
## log(p\_wisk64) + week, data = dt)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.9041 -0.4030 0.0038 0.4226 2.9182   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.4729537 0.1646144 -15.023 < 2e-16 \*\*\*  
## log(p\_tide128) 1.0020010 0.0631073 15.878 < 2e-16 \*\*\*  
## log(p\_tide64) -3.7314302 0.0604782 -61.699 < 2e-16 \*\*\*  
## log(p\_wisk64) 0.3454653 0.0523267 6.602 4.19e-11 \*\*\*  
## week -0.0067235 0.0001186 -56.682 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7056 on 14740 degrees of freedom  
## Multiple R-squared: 0.3611, Adjusted R-squared: 0.3609   
## F-statistic: 2083 on 4 and 14740 DF, p-value: < 2.2e-16

lm128B\_coef<- summary(lm128\_B)$coefficients  
lm64B\_coef<-summary(lm64\_B)$coefficients#Lets graph these coefficients so we can see all these at a glance  
  
pe\_graphB<-c(lm64B\_coef[2:4,1],lm128B\_coef[2:4,1])  
pe\_graphB

## log(p\_tide128) log(p\_tide64) log(p\_wisk64) log(p\_tide128) log(p\_tide64)   
## 1.0020010 -3.7314302 0.3454653 -4.7716627 0.2934515   
## log(p\_wisk64)   
## 0.6294800

barplot(pe\_graphB, main = "Price Elasticity", xlab = "Tide64's PE\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Tide128's PE", col = rainbow(6), names.arg = c('Cross Tide','Own', 'Cross Wisk','Own', 'Cross Tide', 'Cross Wisk' ))



# B

Discussion: Explain why adding a time trend is important here. Discuss whether the demand estimates now make sense. Is there an improvement over the model specification in question 2?

The demand estimates have changed, but only slightly. Not enough to now make sense.

Time trend is extremely important, once we accounted for seasonal trends, it improves the previous model immensely. This makes sense, since the purchase of detergent is likely to be driven by weather(depends on the location, but think about people spending alot of time outside vs inside would change the rate at which the product is used). It is important not to try to build too many assumptions into this. Tide128 is a product that may take a few weeks to consume, and so if there is a seasonal effect driving its consumption there will be a lag effect. The big takeaway is that incorporating the week that the detergent was purchased into our model we get a more accurate model.

See the table below for a comparison of the multiple R squared values. Once we implemented dummy variables for time, our model was able to capture between roughly 8-10% more of the variation.

|  |  |  |  |
| --- | --- | --- | --- |
| Tide64 Linear-Log | Tide64 w/Time | Tide128Linear-Log | Tide 128 w/Time |
| Adj.R^2 0.2217 | 0.3609 | 0.266 | 0.2898 |

## Question 4

# A

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':  
##   
## between, first, last

## The following object is masked from 'package:MASS':  
##   
## select

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

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

detergent\_promos <- dt %>% select(week, promoflag) %>%   
 filter(promoflag == 1)  
  
totalweeks<-length(unique(dt$week))  
  
promoweeks<-length(unique(detergent\_promos$week))  
  
store\_week\_fraction <- paste(promoweeks,totalweeks, sep = '/')  
  
store\_week\_fraction #Did we lose some data, there isn't the expected 300 weeks.

## [1] "195/224"

sum(dt$promoflag)

## [1] 12069

detergent\_DF\_2 = subset(detergent\_DF, promoflag != 1)

# B

detergent\_DF\_2$svelocity128 <- (detergent\_DF\_2$q\_tide128/detergent\_DF\_2$acv)  
detergent\_DF\_2$svelocity64 <- (detergent\_DF\_2$q\_tide64/detergent\_DF\_2$acv)  
  
SalesVelocity128\_B <- sum(detergent\_DF\_2$svelocity128)  
SalesVelocity64\_B <- sum(detergent\_DF\_2$svelocity64)  
  
lm128\_C = lm(log(svelocity128) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64) + week, data = detergent\_DF\_2)  
lm64\_C = lm(log(svelocity64) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64) + week, data = detergent\_DF\_2)  
  
summary(lm128\_C)

##   
## Call:  
## lm(formula = log(svelocity128) ~ log(p\_tide128) + log(p\_tide64) +   
## log(p\_wisk64) + week, data = detergent\_DF\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.6041 -0.3453 0.0227 0.3875 2.1183   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.6394373 0.4112965 1.555 0.120   
## log(p\_tide128) -3.5000136 0.1889015 -18.528 < 2e-16 \*\*\*  
## log(p\_tide64) -0.1158488 0.1480837 -0.782 0.434   
## log(p\_wisk64) 0.7141957 0.1651009 4.326 1.58e-05 \*\*\*  
## week -0.0013283 0.0002534 -5.242 1.71e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6601 on 2671 degrees of freedom  
## Multiple R-squared: 0.1225, Adjusted R-squared: 0.1212   
## F-statistic: 93.21 on 4 and 2671 DF, p-value: < 2.2e-16

summary(lm64\_C)

##   
## Call:  
## lm(formula = log(svelocity64) ~ log(p\_tide128) + log(p\_tide64) +   
## log(p\_wisk64) + week, data = detergent\_DF\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4380 -0.3534 -0.0117 0.3142 3.1648   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.8816576 0.4184072 -6.887 7.07e-12 \*\*\*  
## log(p\_tide128) 0.2675906 0.1921673 1.392 0.16389   
## log(p\_tide64) -1.7018134 0.1506438 -11.297 < 2e-16 \*\*\*  
## log(p\_wisk64) -0.5189392 0.1679553 -3.090 0.00202 \*\*   
## week -0.0069032 0.0002578 -26.782 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6715 on 2671 degrees of freedom  
## Multiple R-squared: 0.3071, Adjusted R-squared: 0.3061   
## F-statistic: 296 on 4 and 2671 DF, p-value: < 2.2e-16

lm128\_Ca<-tidy(lm128\_C)  
lm64\_Ca<-tidy(lm64\_C)  
  
kable(lm128\_Ca[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | 0.6394 | 0.4113 | 1.5547 | 0.1201 |
| log(p\_tide128) | -3.5000 | 0.1889 | -18.5282 | 0.0000 |
| log(p\_tide64) | -0.1158 | 0.1481 | -0.7823 | 0.4341 |
| log(p\_wisk64) | 0.7142 | 0.1651 | 4.3258 | 0.0000 |

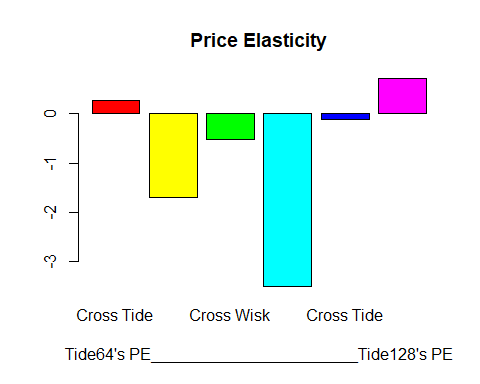
kable(lm64\_Ca[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -2.8817 | 0.4184 | -6.8872 | 0.0000 |
| log(p\_tide128) | 0.2676 | 0.1922 | 1.3925 | 0.1639 |
| log(p\_tide64) | -1.7018 | 0.1506 | -11.2969 | 0.0000 |
| log(p\_wisk64) | -0.5189 | 0.1680 | -3.0897 | 0.0020 |

lm128C\_coef<- summary(lm128\_C)$coefficients  
lm64C\_coef<-summary(lm64\_C)$coefficients#Lets graph these coefficients so we can see all these at a glance  
  
pe\_graphC<-c(lm64C\_coef[2:4,1],lm128C\_coef[2:4,1])  
pe\_graphC

## log(p\_tide128) log(p\_tide64) log(p\_wisk64) log(p\_tide128) log(p\_tide64)   
## 0.2675906 -1.7018134 -0.5189392 -3.5000136 -0.1158488   
## log(p\_wisk64)   
## 0.7141957

barplot(pe\_graphC, main = "Price Elasticity", xlab = "Tide64's PE\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Tide128's PE", col = rainbow(6), names.arg = c('Cross Tide','Own', 'Cross Wisk','Own', 'Cross Tide', 'Cross Wisk' ))



# Discussion

Discuss whether the demand estimates (own and cross price elasticities) now make sense - is there an improvement over the specification in question 3? Provide some intuition for the change in the estimated own-price effects.

BEFORE: OwnPE\_128 = -4.59 OwnPE\_64 = -3.74

CrossPE\_128 = 0.31 CrossPE\_64 = 1.28 Cross\_PE\_128 & Wisk = 0.15 Cross\_PE\_64 & Wisk = -0.87

AFTER: OwnPE\_128 = -3.5 OwnPE\_64 = -1.7018

CrossPE\_128 = 2.676 CrossPE\_64 = -0.1158 Cross\_PE\_128 & Wisk = 0.7142 Cross\_PE\_64 & Wisk = -0.5189

The own price effects have decreased significantly, and also the gap between the two have increased as well. Now, Tide 128 is highly elastic while Tide 64 is mildly elastic. Additionally, as Tide 128 demand increases, prices for Tide 64 decreases and price for Wisk 64 increases. As Tide 64 demand increase, prices for Tide 128 increases and prices for Wisk decreases.It is still interesting that if demand increases for either Tide product, the prices for the opposite Tide product and for the Wisk product are not sharing the same sign (go in opposite directions).

## BEST GUESS?

## Question 5

Store Fixed Effect

# A

lm128\_D<-lm(log(svelocity128) ~ log(p\_tide128)+log(p\_tide64) + log(p\_wisk64) + week + factor(store), data = detergent\_DF\_2)  
lm64\_D <- lm(log(svelocity64) ~ log(p\_tide128)+log(p\_tide64)+ log(p\_wisk64) + week+ factor(store), data = detergent\_DF\_2)  
lm128D\_coef<- summary(lm128\_D)$coefficients  
lm64D\_coef<-summary(lm64\_D)$coefficients#Lets graph these coefficients so we can see all these at a glance  
  
library(broom)  
library(knitr)  
  
lm128\_D <- tidy(lm128\_D)  
lm64\_D <- tidy(lm64\_D)  
  
  
  
kable(lm64\_D[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -4.5977 | 0.6313 | -7.2823 | 0.0000 |
| log(p\_tide128) | 0.9028 | 0.2202 | 4.1005 | 0.0000 |
| log(p\_tide64) | -1.4867 | 0.1635 | -9.0939 | 0.0000 |
| log(p\_wisk64) | -0.2809 | 0.1733 | -1.6205 | 0.1052 |

#Comparison  
kable(lm64\_Ca[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -2.8817 | 0.4184 | -6.8872 | 0.0000 |
| log(p\_tide128) | 0.2676 | 0.1922 | 1.3925 | 0.1639 |
| log(p\_tide64) | -1.7018 | 0.1506 | -11.2969 | 0.0000 |
| log(p\_wisk64) | -0.5189 | 0.1680 | -3.0897 | 0.0020 |

#Comparison  
kable(lm128\_D[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -2.3519 | 0.5089 | -4.6215 | 0.0000 |
| log(p\_tide128) | -2.3836 | 0.1775 | -13.4310 | 0.0000 |
| log(p\_tide64) | 0.2097 | 0.1318 | 1.5912 | 0.1117 |
| log(p\_wisk64) | 1.1648 | 0.1397 | 8.3378 | 0.0000 |

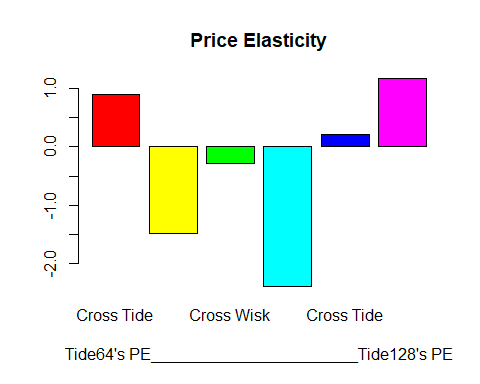
kable(lm128\_Ca[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | 0.6394 | 0.4113 | 1.5547 | 0.1201 |
| log(p\_tide128) | -3.5000 | 0.1889 | -18.5282 | 0.0000 |
| log(p\_tide64) | -0.1158 | 0.1481 | -0.7823 | 0.4341 |
| log(p\_wisk64) | 0.7142 | 0.1651 | 4.3258 | 0.0000 |

pe\_graphD<-c(lm64D\_coef[2:4,1],lm128D\_coef[2:4,1])  
pe\_graphD

## log(p\_tide128) log(p\_tide64) log(p\_wisk64) log(p\_tide128) log(p\_tide64)   
## 0.9028341 -1.4866560 -0.2808741 -2.3836409 0.2096755   
## log(p\_wisk64)   
## 1.1648454

barplot(pe\_graphD, main = "Price Elasticity", xlab = "Tide64's PE\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Tide128's PE", col = rainbow(6), names.arg = c('Cross Tide','Own', 'Cross Wisk','Own', 'Cross Tide', 'Cross Wisk' ))



# B

Discussion: Do the estimates of own and cross price elasticties reveal an improvement over the model specification in question 4?

Yes, the own and cross price elasticities decrease from Question 4’s model to this model for Tide 64. For Tide 128, it’s own price elasticity decreases and more noteably, the cross price elasticity for tide 64 changes sign from negative to positive, which finally makes more sense. This means that as the demand for Tide 128 increase, the price for Tide 64 increases only slightly and for Wisk 64 increases more significantly while the price for Tide 128 decreases. As the demand increases for Tide 64, the price for Tide 128 increases slightly while prices decrease slightly for Wisk 64 and decreases more for Tide 64. This makes more sense beceause if there are price changes for a 64 size product, people might feel like they’re getting a better deal for Tide 128 which is double the size of Tide 64 but now “cheaper”.

# C

lm128\_E<-lm(log(q\_tide128) ~ log(p\_tide128)+log(p\_tide64) +log(p\_wisk64) + week + factor(store), data = detergent\_DF\_2)  
lm64\_E<- lm(log(q\_tide64) ~ log(p\_tide128)+log(p\_tide64) +log(p\_wisk64) + week+ factor(store), data = detergent\_DF\_2)  
  
lm128\_Ea = tidy(lm128\_E)  
lm64\_Ea = tidy(lm64\_E)  
  
kable(lm128\_D[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -2.3519 | 0.5089 | -4.6215 | 0.0000 |
| log(p\_tide128) | -2.3836 | 0.1775 | -13.4310 | 0.0000 |
| log(p\_tide64) | 0.2097 | 0.1318 | 1.5912 | 0.1117 |
| log(p\_wisk64) | 1.1648 | 0.1397 | 8.3378 | 0.0000 |

kable(lm128\_Ea[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | 7.1826 | 0.5089 | 14.1140 | 0.0000 |
| log(p\_tide128) | -2.3836 | 0.1775 | -13.4310 | 0.0000 |
| log(p\_tide64) | 0.2097 | 0.1318 | 1.5912 | 0.1117 |
| log(p\_wisk64) | 1.1648 | 0.1397 | 8.3378 | 0.0000 |

kable(lm64\_D[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -4.5977 | 0.6313 | -7.2823 | 0.0000 |
| log(p\_tide128) | 0.9028 | 0.2202 | 4.1005 | 0.0000 |
| log(p\_tide64) | -1.4867 | 0.1635 | -9.0939 | 0.0000 |
| log(p\_wisk64) | -0.2809 | 0.1733 | -1.6205 | 0.1052 |

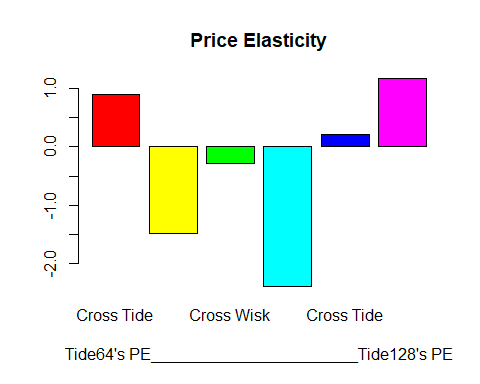
kable(lm64\_Ea[c(1:4),], digits = 4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | 4.9368 | 0.6313 | 7.8195 | 0.0000 |
| log(p\_tide128) | 0.9028 | 0.2202 | 4.1005 | 0.0000 |
| log(p\_tide64) | -1.4867 | 0.1635 | -9.0939 | 0.0000 |
| log(p\_wisk64) | -0.2809 | 0.1733 | -1.6205 | 0.1052 |

lm128E\_coef<- summary(lm128\_E)$coefficients  
lm64E\_coef<-summary(lm64\_E)$coefficients#Lets graph these coefficients so we can see all these at a glance  
  
pe\_graphE<-c(lm64E\_coef[2:4,1],lm128E\_coef[2:4,1])  
pe\_graphE

## log(p\_tide128) log(p\_tide64) log(p\_wisk64) log(p\_tide128) log(p\_tide64)   
## 0.9028341 -1.4866560 -0.2808741 -2.3836409 0.2096755   
## log(p\_wisk64)   
## 1.1648454

barplot(pe\_graphE, main = "Price Elasticity", xlab = "Tide64's PE\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Tide128's PE", col = rainbow(6), names.arg = c('Cross Tide','Own', 'Cross Wisk','Own', 'Cross Tide', 'Cross Wisk' ))



Discussion: How do the elasticity estimates and the time trend compare across these two regressions? Is the difference (or absence of a difference) as expected?

They are surprisingly identical, using unit sales instead of velocity did not have any effect. This doesn’t make sense conceptually, large stores and small stores should have variation between them. When we removed our velocity it did nothing to change the elasticities.

## Question 6

r = 25% C = 0.25 / oz

Q128 = A128 \* P128^(-2.3836) \* P64^(0.2097) \* PW64^(1.1648) Q64 = A64 \* P128^(0.9028) \* P64^(-1.4867) \* PW64^(-0.2809)

# A

baseprice128 = mean(detergent\_DF\_2$p\_tide128)  
baseprice128

## [1] 8.474708

# B

baseprice64 = mean(detergent\_DF\_2$p\_tide64)  
baseprice64

## [1] 4.399403

# C / D

stores <- length(unique(detergent\_DF\_2$store))  
  
basevolume128 = stores \* 52 \* mean(detergent\_DF\_2$q\_tide128)  
basevolume64 = stores \* 52 \* mean(detergent\_DF\_2$q\_tide64)  
  
basevolume128

## [1] 247068

basevolume64

## [1] 282954.3

# E

Base Profits: pi128=Q\_128\*(P\_128(1-r)-C)

Bprofit128 = basevolume128 \* ((baseprice128\*0.75) - 0.25)  
Bprofit64 = basevolume64 \* ((baseprice64\*0.75) - 0.25)  
  
BaseTotalProfit <- Bprofit128 + Bprofit64  
BaseTotalProfit

## [1] 2371489

# F

Scenario1: 5% increase in price for both 128 and 64 Scenario2: 5% decrease in price for both 128 and 64 Scenario3: 5% increase in price 128 and 5% decrease in 64 Scenario4: 5% decrease in price 128 and 5% increase in 64

r = 25% C = 0.25 / oz

Q128 = A128 \* P128^(-2.3836) \* P64^(0.2097) \* PW64^(1.1648) Q64 = A64 \* P128^(0.9028) \* P64^(-1.4867) \* PW64^(-0.2809)

pricechange = 0.05  
  
# Specify the function  
demand\_change <- function(price\_change1, price\_change2, price\_elasticity1, price\_elasticity2, price\_elasticity3, basequantity) {  
   
 # New demand from method 1   
 quantity\_change\_1 = ((1 + price\_change1)^price\_elasticity1) \* ((1 + price\_change2)^price\_elasticity2) \*  
 ((1)^price\_elasticity3)  
   
 #Get new quantity from ratio of new / old   
   
 NewQuantity = quantity\_change\_1 \* basequantity  
   
 # Return results as a list  
 return(NewQuantity)  
}

Running the Function

New\_Q128\_1 <- demand\_change(0.05, 0.05, -2.11, 0.211,1.1648, basevolume128)  
New\_Q64\_1 <- demand\_change(0.05,0.05,1.16,-1.585, -0.2809, basevolume64)  
  
New\_Q128\_2 <- demand\_change(-0.05, -0.05, -2.11, 0.211, 1.1648, basevolume128)  
New\_Q64\_2 <- demand\_change(-0.05,-0.05,1.16,-1.585 , -0.2809, basevolume64)  
  
New\_Q128\_3 <- demand\_change(0.05, 0.05, -2.11, 0.211, 1.1648, basevolume128)  
New\_Q64\_3 <- demand\_change(-0.05,-0.05,1.16,-1.585 , -0.2809, basevolume64)  
  
New\_Q128\_4 <- demand\_change(-0.05, -0.05, -2.11, 0.211, 1.1648, basevolume128)  
New\_Q64\_4 <- demand\_change(0.05,0.05,1.16,-1.585 , -0.2809, basevolume64)

# G

#Total Profit calculations  
#TotalProfit1 <- (New\_Q128\_1\* (baseprice128\*1.05))\*(New\_Q64\_1\* (baseprice64\*1.05))  
  
#TotalProfit2 <- (New\_Q128\_2\*(baseprice128\*0.95))\*(New\_Q64\_2\* (baseprice64\*0.95))  
  
#TotalProfit3 <- (New\_Q128\_3\*(baseprice128\*1.05))\*(New\_Q64\_3\*(baseprice64\*0.95))  
  
#TotalProfit4 <- (New\_Q128\_4\*(baseprice128\*0.95))\*(New\_Q64\_4\* (baseprice64\*1.05))

### CORRECT PROFIT?

Profit Equation: pi128=Q\_128*(P\_128(1-r)-C) pi64 = Q\_64*(P\_64(1-r)-C) r = 25% C = 0.25 / oz

Profit1 = (New\_Q128\_1 \* ((baseprice128\*1.05\*0.75) - 0.25)) + (New\_Q64\_1 \* ((baseprice64\*1.05\*0.75) - 0.25))  
Profit2 = (New\_Q128\_2 \* ((baseprice128\*0.95\*0.75) - 0.25)) + (New\_Q64\_2 \* ((baseprice64\*0.95\*0.75) - 0.25))  
Profit3 = (New\_Q128\_3 \* ((baseprice128\*1.05\*0.75) - 0.25)) + (New\_Q64\_3 \* ((baseprice64\*0.95\*0.75) - 0.25))  
Profit4 = (New\_Q128\_4 \* ((baseprice128\*0.95\*0.75) - 0.25)) + (New\_Q64\_4 \* ((baseprice64\*0.95\*0.75) - 0.25))

: Table of quantities sold and profits when Tide changes the price of Tide 64 and 128. Price changes are shown in percentages

library(scales)

## Warning: package 'scales' was built under R version 4.0.3

##   
## Attaching package: 'scales'

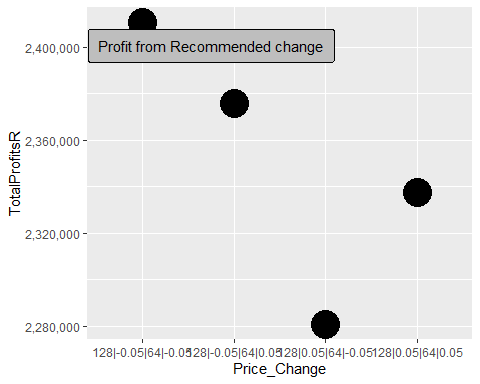
## The following object is masked from 'package:arm':  
##   
## rescale

## The following objects are masked from 'package:psych':  
##   
## alpha, rescale

del\_price\_128 = c(0.05, -0.05, 0.05, -0.05)  
del\_price\_64 = c(0.05, -0.05, -0.05, 0.05)  
q\_128 = c(New\_Q128\_1, New\_Q128\_2, New\_Q128\_3, New\_Q128\_4)  
q\_64 = c(New\_Q64\_1, New\_Q64\_2, New\_Q64\_3, New\_Q64\_4)  
#TotalProfitsW = c(TotalProfit1, TotalProfit2, TotalProfit3, TotalProfit4)  
TotalProfitsR = c(Profit1, Profit2, Profit3, Profit4)  
  
Price\_Change<- as.factor(paste('128',del\_price\_128,'64',del\_price\_64, sep = "|"))  
  
NewTable <- data.frame(del\_price\_128, del\_price\_64, q\_128, q\_64,Price\_Change, TotalProfitsR)  
  
NewTable

## del\_price\_128 del\_price\_64 q\_128 q\_64 Price\_Change TotalProfitsR  
## 1 0.05 0.05 225205.0 277147.4 128|0.05|64|0.05 2337578  
## 2 -0.05 -0.05 272344.9 289190.3 128|-0.05|64|-0.05 2410586  
## 3 0.05 -0.05 225205.0 289190.3 128|0.05|64|-0.05 2280870  
## 4 -0.05 0.05 272344.9 277147.4 128|-0.05|64|0.05 2375848

ggplot(data = NewTable, aes(Price\_Change,TotalProfitsR))+geom\_point(size=10,show.legend = NA)+scale\_y\_continuous(labels = comma)+geom\_label(  
 label="Profit from Recommended change",   
 x=1.75,  
 y=2400500,  
 label.padding = unit(0.55, "lines"), # Rectangle size around label  
 label.size = 0.35,  
 color = "black",  
 fill="grey")



## Question 7

1. What is the extent of cannibalization within the Tide product line?

Q128 = A128 \* P128^(-2.3836) \* P64^(0.2097) \* PW64^(1.1648) Q64 = A64 \* P128^(0.9028) \* P64^(-1.4867) \* PW64^(-0.2809)

Cross PE of 128 is 0.9028 which is > Cross PE of 64 at 0.2097.

This suggests that Tide 128 cannibalizes Tide 64.

1. Does Tide face a competitive threat from Wisk?

A little bit. If Tide 128 price goes up and its demand decreases, price for Tide 64 goes up slightly but price for Wisk goes up more. However, if Tide 64 price increases and its demand decreases, Tide 128 goes up while Wisk price decreases but only slightly. There is competition for Tide 128, but not Tide 64.

1. How do you evaluate the current pricing tactics? Do you recommend changes? Base Profit: 2371489 If you decrease both Tide product’s prices by 5%, the expected profit is 2410586. This is therefore our recommendation, since it is > Base Profit.