Group Project Report - Sports Drinks

SCM 651

Fall 2017

Prof. Basu

Submitted by

Siyao Xu

Yifan Liu

Pan Chen

I. Executive Summary:

The general purpose of our final project is to study the possible factors that affect Dominick's grocery store chains' sales of different brands of sports drinks from 9/14/89 to 5/8/97 in the Chicago area. We are also trying to find out if different factors attribute to the different sale figures of different brands. All of the data preparation and analysis were done by RStudio, its base package, and "xlsx" package.

Our dataset included the sales data of the various products from three sports drink brands: Gatorade, Powerade, and Allsports, and then we decided to focus on the best-seller from each brand.

Here are our conclusions:

Among the three brands, Gatorade dominated the sales, which is not too surprising considering year 1989-1997 was the prime era of Chicago Bulls and Gatorade is the official sports drink of NBA, and the brand Bulls' star Michael Jordan represented; however, Powerade's sale is the poorest, which is surprising because Powerade is manufactured by Coca Cola Company and it could not even outsell lesser-known brand Allsport. Powerade is also more expensive than the other two brands on average at 32oz. In addition, Powerade has the highest price elasticity of demand comparing with the other two brands, which means if the three brands reduce the same amount of the price, Powerade's sale will be boosted more significantly than the other two. We believe that the higher price and its relatively tendency to be on sale could be the reasons contributing to its lower market share than its competitors, and by lowering its price and being put on sale more might gain Powerade some competitive edge.

On the other hand, for the store brand like Allsport to compete with the national brands, store brands like Allsport could be promoted more in the areas with more senior people, minorities and unemployed people.

Since Gatorade absolutely dominated the sales, we do not have much suggestions for the improvement of its sales.

The following sections demonstrated how we reached the conclusion in the executive summary:

II.   Dataset Preparation:

The datasets we used for this project are weekly movement data for high movement sports drinks (sports drinks high movement.csv), limited version of store demographics data (DEMOSMALL.xlsx), and UPC codes and product descriptions for high movement items(sports drinks upc.xlsx).

To compile the complete dataset that we need for the analysis, we first merged the weekly movement data and store demographics data by "STORE" column; then we merged the merged dataset from the previous step with UPC data by "UPC" column.

Then we decided to focus our analysis on the best-seller of each brand, because we thought they might be the better indicators of the effects of various factors we want to analyze. To do so, we have aggregated the sum of MOVE of each UPC, and the results are as follows (the orange columns highlight Gatorade products, the blue columns highlight All Sport products, and the grey columns highlight Powerade products):

UPCs with Most Moves
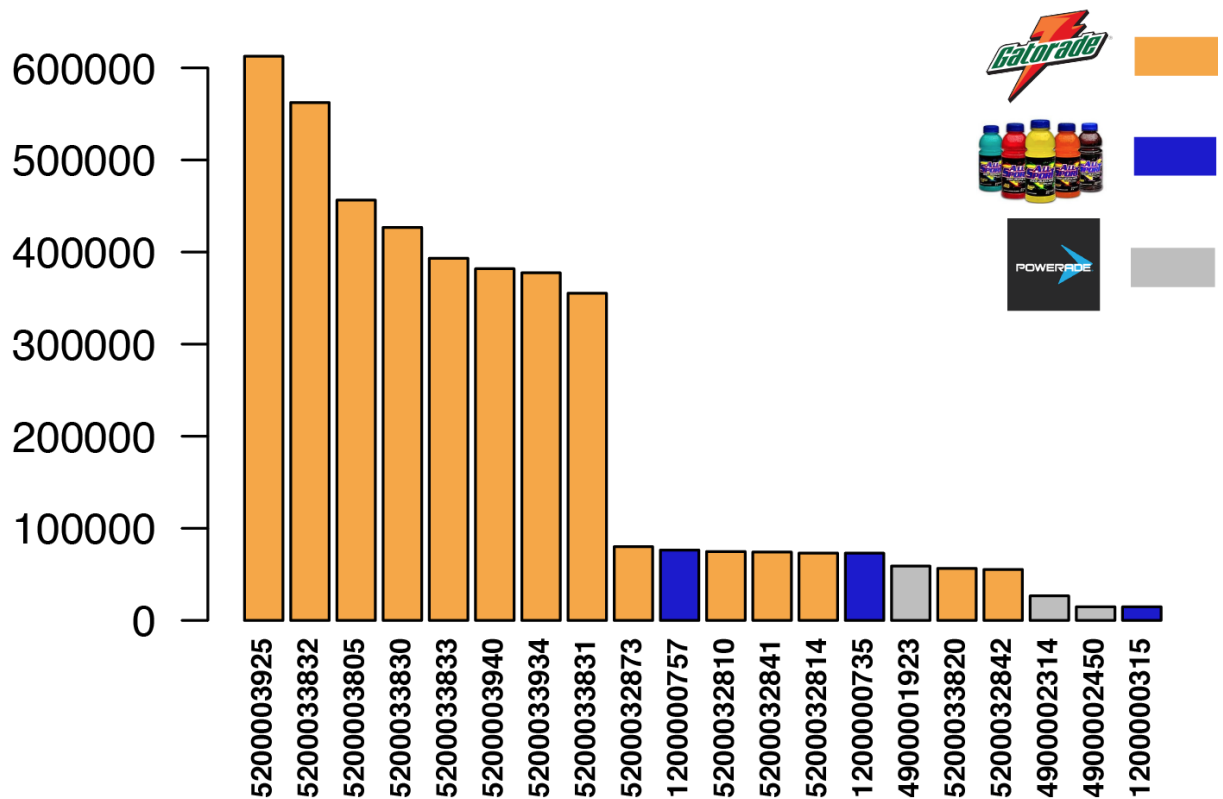
As we can see, Gatorade products dominated the graph in terms of units moved, and Powerade, despite coming from a mega-company Coca Cola, were surprisingly outsold by less-known All Sport Drink.

Based on our criteria, we have selected these three products for our analysis, which are the best-sellers of each brand respectively:

UPC	Brand	Variety	Brand Type
5200003925	GATORADE	LEMON/LIME 32OZ 12	National
4900001923	POWERADE	FRUIT PUNCH 32OZ 12	National
1200000757	ALL SPORT	FRUIT PUNC 32OZ 12	Less-known

After that we filtered other UPC out of our dataset, and therefore achieved the final dataset we are going to use for our data analysis.

Our final dataset include these columns:UPC, STORE, STOREWEEK, WEEK, BRAND, SIZE, MOVE, logmove, QTY, PRICE, logprice, PROFIT, Feat, NAME, CITY, ZIP, LAT, LONG, AGE9, AGE60, ETHNIC, EDUC, NOCAR, INCOME, HHSINGLE, HHLARGE, WORKWOM, HVAL150, SINGLE, RETIRED, UNEMP, NWHITE, POVERTY, DRTIME5, SSTRDIST, SSTRVOL, CPDIST5, CPWVOL5, PRICLOW, PRICMED, PRICHIGH, CASE. Please refer to the project guideline to see the descriptions of each column.

III. Research Questions:

1. How does the price vary across brands?

2. What factors affect demand of national brands and store brands?

3. What factors affect demand of national brands?

4. What is the price elasticity of demand of a brand?

5. Is price elasticity different for different brands?

6. How does the probability that the product is on sale depend on brand?

IV. Findings

1. How does the price vary across brands?

We have calculated the average price of each brand's best-seller (they are all in 32oz bottles). The results are as follows:

Average price of 32oz Bottle of Each Brand
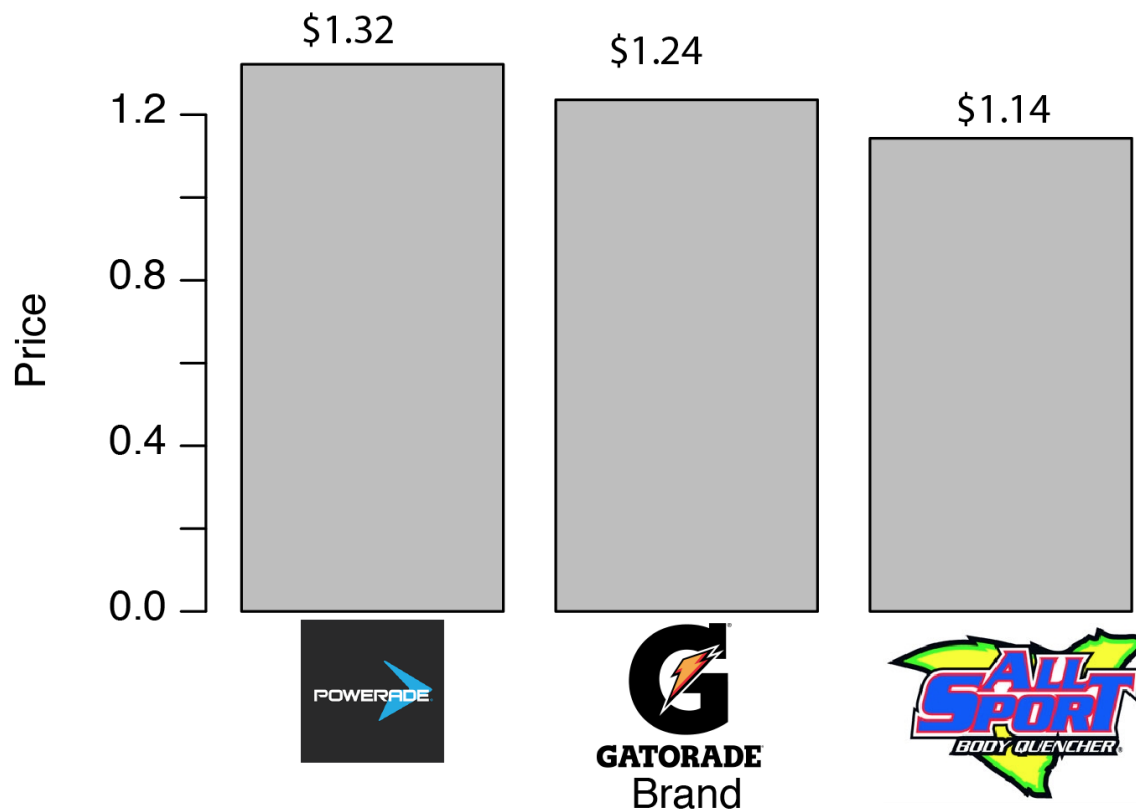
Of the best-seller we chose from each brand, it looks like Powerade is the most expensive one, with $1.32 per 32oz bottle, and All Sport is the least expensive one, with $1.14 per 32oz bottle. Gatorade is in the middle, with $1.24 per 32oz bottle. This is consistent with the brands' target audiences, since less-known brands like All Sport often set lower price to compete with national brands.

2. **What factors affect demand of national brands and store brands?**

National Brands (Gatorade Lemon/Lime and Powerade Fruit Punch) Results:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.73939    1.14918  -2.384   0.0171 *
logprice    -3.53513    0.03845 -91.953  < 2e-16 ***
Feat         0.02439    0.01336   1.825   0.0680 .
AGE9        10.09375    1.03416   9.760  < 2e-16 ***
AGE60        1.24604    0.64542   1.931   0.0535 .
ETHNIC       0.24656    0.18476   1.334   0.1821
EDUC        -0.80743    0.14460  -5.584 2.37e-08 ***
NOCAR        1.51259    0.18781   8.054 8.28e-16 ***
INCOME       0.61710    0.08551   7.217 5.43e-13 ***
HHSINGLE    -1.93042    0.33696  -5.729 1.02e-08 ***
HHLARGE     -4.12840    0.61533  -6.709 1.99e-11 ***
WORKWOM     -4.10954    0.65520  -6.272 3.60e-10 ***
HVAL150      0.46186    0.06497   7.108 1.20e-12 ***
SINGLE       5.07127    0.42732  11.867  < 2e-16 ***
RETIRED     -1.35289    0.81132  -1.668   0.0954 .
UNEMP       -5.69401    1.22918  -4.632 3.63e-06 ***
NWHITE      -1.36933    0.17593  -7.783 7.25e-15 ***
POVERTY     -0.32130    0.73271  -0.439   0.6610
```

Store Brand (All Sport Fruit Punch) Results:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -18.02062    2.51353  -7.169 8.54e-13 ***
logprice     -2.30831    0.11859 -19.464  < 2e-16 ***
Feat          0.26628    0.02919   9.124  < 2e-16 ***
AGE9          5.48206    2.31767   2.365  0.01805 *
AGE60         9.26076    1.43680   6.445 1.25e-10 ***
ETHNIC        2.14863    0.45798   4.692 2.78e-06 ***
EDUC         -0.90665    0.31926  -2.840  0.00453 **
NOCAR         1.79979    0.44696   4.027 5.73e-05 ***
INCOME        1.51009    0.18668   8.089 7.33e-16 ***
HHSINGLE     -1.63672    0.78071  -2.096  0.03609 *
HHLARGE      -4.30476    1.37220  -3.137  0.00172 **
WORKWOM       3.06154    1.42588   2.147  0.03183 *
HVAL150      -0.43015    0.14295  -3.009  0.00263 **
SINGLE        1.26155    0.96382   1.309  0.19062
RETIRED      -7.10136    1.80215  -3.940 8.23e-05 ***
UNEMP        15.23065    2.74343   5.552 2.96e-08 ***
NWHITE       -2.56175    0.44788  -5.720 1.12e-08 ***
POVERTY      -0.16156    1.68765  -0.096  0.92374
```

To compare the differences between national and store brands, we did a analysis based on the two datasets: national brands and store brand. Based on the outputs above, we concluded that some demographic variables such as **age60, ethnic, retired, and single** do not have the same effects on both national and store brands. For instance, **SINGLE** significantly affects the sale of national brands but not store brand.

3. What factors affect demand of national brands?

<u>Gatorade Results</u>

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.42480    1.15631  -3.827  0.00013 ***
logprice    -2.58136    0.04370 -59.069  < 2e-16 ***
Feat         0.20724    0.01365  15.178  < 2e-16 ***
AGE9        11.89278    1.03810  11.456  < 2e-16 ***
AGE60        0.67809    0.64983   1.043  0.29673
ETHNIC       0.30948    0.18480   1.675  0.09400 .
EDUC        -0.76879    0.14535  -5.289 1.24e-07 ***
NOCAR        1.30930    0.18938   6.914 4.83e-12 ***
INCOME       0.78609    0.08596   9.145  < 2e-16 ***
HHSINGLE    -2.71725    0.33695  -8.064 7.66e-16 ***
HHLARGE     -5.12756    0.61912  -8.282  < 2e-16 ***
WORKWOM     -4.42895    0.65910  -6.720 1.86e-11 ***
HVAL150      0.38211    0.06566   5.820 5.96e-09 ***
SINGLE       6.50456    0.42914  15.157  < 2e-16 ***
RETIRED     -0.29627    0.81558  -0.363  0.71641
UNEMP       -7.97341    1.23576  -6.452 1.12e-10 ***
NWHITE      -1.90967    0.17542 -10.886  < 2e-16 ***
POVERTY      1.77148    0.73973   2.395  0.01664 *
```

<u>Powerade Results</u>

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.99562    2.31663  -0.861 0.389029
logprice     -2.98648    0.06847 -43.615  < 2e-16 ***
Feat          0.38989    0.02878  13.549  < 2e-16 ***
AGE9          4.49865    2.11155   2.131 0.033166 *
AGE60         4.55035    1.29900   3.503 0.000463 ***
ETHNIC        0.39271    0.38174   1.029 0.303636
EDUC         -0.56166    0.29234  -1.921 0.054739 .
NOCAR         2.03764    0.37509   5.432 5.75e-08 ***
INCOME        0.42563    0.17313   2.458 0.013981 *
HHSINGLE     -2.43897    0.69664  -3.501 0.000466 ***
HHLARGE      -5.02958    1.24013  -4.056 5.05e-05 ***
WORKWOM      -1.72151    1.32665  -1.298 0.194456
HVAL150       0.24158    0.12974   1.862 0.062645 .
SINGLE        0.67944    0.86838   0.782 0.433994
RETIRED      -8.05021    1.65029  -4.878 1.10e-06 ***
UNEMP         4.29162    2.49363   1.721 0.085290 .
NWHITE       -0.95833    0.36998  -2.590 0.009612 **
POVERTY       1.45579    1.46716   0.992 0.321112
```

To further analyze the data, we broke down the national brands dataset into two subsets based on brand: Gatorade and Powerade. For Gatorade, demographic variables such as age60, ethnic, and retired do not have a significant effect on the demand of Gatorade sport drinks. In contrast, ethnic, education, workwom, hval150, single and unemployment are deemed insignificant at a 95% level of confidence.

4. What is the price elasticity of demand of a brand?

We computed the price elasticity of demand of each brand based on the linear model results:

Brand	Gatorade	Powerade Fruit	All Sport Fruit

	Lemon/Lime	Punch	Punch
Price Elasticity of Demand	**2.50428**	**3.20761**	**2.34444**

5. Is price elasticity different for different brands?

The price elasticity of demand is different for each brand. Among the three brands, demand of Powerade sport drinks is the most sensitive to a price change, while demand of All Sport product is the least sensitive to change in price.

<u>Liner Model Results</u>

```
lm(formula = logmove ~ logprice + Feat + BRAND + logprice * BRAND +
    AGE9 + AGE60 + ETHNIC + EDUC + NOCAR + INCOME + HHSINGLE +
    HHLARGE + WORKWOM + HVAL150 + SINGLE + RETIRED + UNEMP +
    NWHITE + POVERTY, data = newdf)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6363 -0.5288  0.0310  0.5488  3.1427

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -6.35462    0.96266  -6.601 4.13e-11 ***
logprice                          -2.34444    0.09896 -23.692  < 2e-16 ***
Feat                               0.24443    0.01141  21.428  < 2e-16 ***
BRANDGATORADE LEMON/LIME           0.69511    0.01990  34.923  < 2e-16 ***
BRANDPOWERADE FRUIT PUNCH         -0.17061    0.02392  -7.133 1.00e-12 ***
AGE9                              10.10218    0.86826  11.635  < 2e-16 ***
AGE60                              2.65335    0.54167   4.898 9.70e-07 ***
ETHNIC                             0.59552    0.15606   3.816 0.000136 ***
EDUC                              -0.76613    0.12125  -6.319 2.67e-10 ***
NOCAR                              1.56060    0.15819   9.866  < 2e-16 ***
INCOME                             0.83649    0.07162  11.679  < 2e-16 ***
HHSINGLE                          -2.50487    0.28416  -8.815  < 2e-16 ***
HHLARGE                           -5.05660    0.51606  -9.798  < 2e-16 ***
WORKWOM                           -3.16672    0.54798  -5.779 7.57e-09 ***
HVAL150                            0.24206    0.05453   4.439 9.07e-06 ***
SINGLE                             4.91221    0.35928  13.672  < 2e-16 ***
RETIRED                           -2.92345    0.68005  -4.299 1.72e-05 ***
UNEMP                             -3.08295    1.03152  -2.989 0.002803 **
NWHITE                            -1.89006    0.14902 -12.684  < 2e-16 ***
POVERTY                            1.58935    0.61773   2.573 0.010089 *
logprice:BRANDGATORADE LEMON/LIME -0.15984    0.10362  -1.543 0.122951
logprice:BRANDPOWERADE FRUIT PUNCH -0.86317   0.10869  -7.941 2.05e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8374 on 40971 degrees of freedom
  (1517 observations deleted due to missingness)
Multiple R-squared:  0.3761,    Adjusted R-squared:  0.3758
F-statistic:  1176 on 21 and 40971 DF,  p-value: < 2.2e-16
```

6. How the probability that the product is on sale (Feat = 1) depends on Brand?

Logit Model

GLModel<- glm(Feat ~ BRAND, family=binomial(logit), data=newdf)

summary(GLModel)

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3665  -0.6482  -0.6482   0.9994   1.8240

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                  0.43421    0.02691   16.13   <2e-16 ***
BRAND[T.GATORADE LEMON/LIME] -1.88755    0.03075  -61.38   <2e-16 ***
BRAND[T.POWERADE FRUIT PUNCH] -1.24603   0.03699  -33.68   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

exp(coef(GLModel)) # Exponentiated coefficients ("odds ratios")

```
> exp(coef(GLModel))  # Exponentiated coefficients ("odds ratios")
                (Intercept)  BRAND[T.GATORADE LEMON/LIME]
                  1.5437363                     0.1514426
BRAND[T.POWERADE FRUIT PUNCH]
                  0.2876452
```

Based on the output produced by R, we figured out the probability that a product is likely to be on sale (Feat=1) is 8.93%, 15.71% and 39.31% for Gatorade Lemon/Lime, Powerade Fruit Punch, and All Sport Fruit Punch, respectively. The calculations are given below:

Equation: I = -0.43421 – 1.88755*Gatorade – 1.24603*Powerade

Gatorade: I=-0.43421-1.88755=-2.32176 → P(Y=1)=1/(1+e^(2.32176)=8.93%

Powerade: I=-0.43421-1.24603=-1.68024 → P(Y=1)=1/(1+e^(1.68024)=15.71%

All Sport: I=-0.43421 → P(Y=1)=1/(1+e^(0.43421)=39.31%

the higher the I is, the more likely it's gonna be on sale.

Among these three products, All Sport Fruit Punch is most likely to be on sale(39.31%), Powerade is second likely to be on sale(15.71%), Gatorade is the least likely to be on sale(8.93%).