Substandard Deviations Analysis

Javier Valcarcel & Justin Shalap

September 20, 2018

Data Wrangling

Load relevant R libraries

```
library(tidyverse)
library(readxl)
library(lubridate)
library(ggplot2)
library(hrbrthemes)
library(gcookbook)
```

```
sales <- read_excel("product_sales.xlsx", sheet = 1)

head(sales)
```

```
## # A tibble: 6 x 4
##   month               company    product  total_sales_USD
##   <dttm>              <chr>      <chr>              <dbl>
## 1 2009-03-01 00:00:00 Voltswitch Nova 1            127411
## 2 2009-03-01 00:00:00 Voltswitch Nova 1 M          138302
## 3 2009-03-01 00:00:00 Voltswitch Nova 2            131839
## 4 2009-03-01 00:00:00 Voltswitch Mycro 1                0
## 5 2009-03-01 00:00:00 Voltswitch Nova 3                 0
## 6 2009-03-01 00:00:00 Voltswitch Mycro 2                0
```

Data transformation of Voltswitch sales

We see the structure of the data and wish to transform it into a format that can be merged with other licensee sales data for comparison and regression. Specifically we want to know their total sales, total sales of products using InfiniCharge and what percentage of their revenue came from infringing technology.

```
mycro_line <- c("Mycro 1", "Mycro 2", "Mycro 3")

total_sales_volt <- sales %>%
  filter(company == "Voltswitch") %>%
  mutate(year = year(month)) %>%
  group_by(year) %>%
  summarise(total_sales = sum(total_sales_USD))

total_sales_volt_mycro <- sales %>%
  filter(product %in% mycro_line) %>%
  mutate(year = year(month)) %>%
  group_by(year) %>%
  summarise(total_sales_mycro  = sum(total_sales_USD))

volt_sales <- merge(total_sales_volt, total_sales_volt_mycro)
```

```
volt_sales1 <- volt_sales %>%
  mutate(percent_patent = total_sales_mycro / total_sales)

head(volt_sales1)
```

```
##   year total_sales total_sales_mycro percent_patent
## 1 2009     4817169            853537      0.1771864
## 2 2010     8626491           2566663      0.2975327
## 3 2011    11949673           4302745      0.3600722
## 4 2012    13539102           5151277      0.3804741
## 5 2013    13848538           4419741      0.3191486
## 6 2014     9490083           2146065      0.2261376
```

Data transformation of licensee sales

From here we load, clean and merge the sales data from the other licensees to merge with the Voltswitch data.

```
companies <- c("Year", "Rekall", "Tyrell Corp.", "Sirius Cybernetics",
               "Cyberdyne Systems", "Nakatomi Corp.")

licensed_sales <- read_excel("sales_data.xlsx", sheet = 1, cell_rows(7:14))

licensed_sales_clean <- licensed_sales[, c(1, 3, 4, 5, 6, 7)]
names(licensed_sales_clean) <- companies

licensed_patent_clean <- licensed_sales[, c(1, 9, 10, 11, 12, 13)]
names(licensed_patent_clean) <- companies

lsc <- licensed_sales_clean %>%
  gather(company, sales, -Year)

lpc <- licensed_patent_clean %>%
  gather(company, percent, -Year)

clean_sales_patents <- merge(lsc, lpc) %>%
  arrange(Year, company)

head(clean_sales_patents)
```

```
##   Year            company     sales   percent
## 1 2008  Cyberdyne Systems  103638.3 0.7368421
## 2 2008     Nakatomi Corp.       0.0 0.0000000
## 3 2008             Rekall  660776.3 0.5188679
## 4 2008 Sirius Cybernetics       0.0 0.0000000
## 5 2008       Tyrell Corp.  382268.2 1.0000000
## 6 2009  Cyberdyne Systems 4718794.4 0.7473684
```

Merge Voltswitch with other licensees

We now are able to merge the Voltswitch data with other licensees to view a combined data set ready for regression.

```r
volt_sales2 <- volt_sales1 %>%
  mutate(Year = year) %>%
  mutate(company = "Voltswitch") %>%
  mutate(sales = total_sales) %>%
  mutate(percent = percent_patent) %>%
  select(Year, company, sales, percent)

total_sales_df <- rbind(clean_sales_patents, volt_sales2) %>% arrange(Year, company)

royalty_schedule <- tibble(
  company = companies[-1],
  royalty_percent = c(0.05, 0.07, 0.02, 0.03, 0.01))

reg_tot1 <- total_sales_df %>%
  merge(royalty_schedule) %>%
  mutate(royalties_due = sales * percent * royalty_percent)

reg_tot2 <- reg_tot1 %>%
  group_by(company) %>%
  summarise(tot_sales = sum(sales),
            avg_percent = mean(percent),
            royalty_percent = mean(royalty_percent))

head(reg_tot2)
```

```
## # A tibble: 5 x 4
##   company            tot_sales avg_percent royalty_percent
##   <chr>                  <dbl>       <dbl>           <dbl>
## 1 Cyberdyne Systems  60708026.       0.746            0.03
## 2 Nakatomi Corp.     85491386.       0.122            0.01
## 3 Rekall              6340256.       0.488            0.05
## 4 Sirius Cybernetics 45640221.       0.202            0.02
## 5 Tyrell Corp.       11729122.       1                0.07
```

Analysis and Modeling

Estimate reasonable royalties

Here we use linear regression to estimate the best fit given the sales data and share of sales by InfiniCharge products. We also calculate the royalties due.

```r
roy_mod <- lm(royalty_percent ~ tot_sales + avg_percent, reg_tot2)

volt_sales3 <- volt_sales2 %>%
  summarise(company = "Voltswitch",
            tot_sales = sum(sales),
            avg_percent = mean(percent))

pred_roy_percent <- predict(roy_mod, volt_sales3)

volt_sales3$royalty_percent <- pred_roy_percent
```

```
reg_tot3 <- rbind(reg_tot2, volt_sales3)

reg_tot3 %>%
  mutate(royalties_due = tot_sales * avg_percent * royalty_percent)
```

```
## # A tibble: 6 x 5
##   company            tot_sales avg_percent royalty_percent royalties_due
##   <chr>                  <dbl>       <dbl>           <dbl>         <dbl>
## 1 Cyberdyne Systems  60708026.       0.746          0.03        1358399.
## 2 Nakatomi Corp.     85491386.       0.122          0.01         104422.
## 3 Rekall              6340256.       0.488          0.05         154661.
## 4 Sirius Cybernetics 45640221.       0.202          0.02         184174.
## 5 Tyrell Corp.       11729122.       1              0.07         821039.
## 6 Voltswitch         62271056        0.293          0.0200       365528.
```

Validation of reasonable royalties to defeat Georgia-Pacific defense.

We look to validate that the estimated royalty given the presence of similar licenses is profitable to demonstrate mutual gains to trade for the license. We should expect the price to be above the cost of the license since there are no close substitutes. We use the split- apply-combine methodology.

```
results <- read_excel("survey_results.xlsx", sheet = "Survey Results", cell_rows(4:54))

key <- read_excel("survey_results.xlsx", sheet = "Question Key", cell_rows((4:28)))

results1 <- results[, -1]

split_prices <- split(results1, seq(1:50))
prices_lists <- lapply(split_prices, t)
prices_lists <- lapply(prices_lists, cbind, key)
tester <- bind_rows(prices_lists)

tester_names <- c("Price", "Question", "Talk_Time", "Weight", "PPI", "Camera")
names(tester) <- tester_names

head(tester)
```

```
##   Price Question Talk_Time Weight PPI Camera
## 1   205        1        10      4 300     12
## 2   185        2        10      4 300      8
## 3   210        3        10      4 400     12
## 4   200        4        10      4 400      8
## 5   185        5        10      5 300     12
## 6   170        6        10      5 300      8
```

```
tester_model <- lm(Price ~ Talk_Time + Weight + PPI + Camera, tester)

tester_model
```

```
##
## Call:
## lm(formula = Price ~ Talk_Time + Weight + PPI + Camera, data = tester)
##
## Coefficients:
```

```
## (Intercept)    Talk_Time       Weight          PPI       Camera
##      3.1750       9.9954     -10.5750       0.2477       5.0062
```

```
anova(tester_model)
```

```
## Analysis of Variance Table
##
## Response: Price
##             Df   Sum Sq  Mean Sq   F value    Pr(>F)
## Talk_Time    1 18649337 18649337 227329.12 < 2.2e-16 ***
## Weight       1    33549    33549    408.95 < 2.2e-16 ***
## PPI          1   184140   184140   2244.61 < 2.2e-16 ***
## Camera       1   120300   120300   1466.42 < 2.2e-16 ***
## Residuals 1195    98034       82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Predicting prices for different products

Given the model for prices based on the survey results, we now use that to price best alternatives to satisfy the but-for process. But-for the infringement what would Voltswitch's alternatives have been? We price those and analyze whether those alternatives were as profitable as using InfiniCharge.

```
desc <- read_excel("product_descriptions.xlsx", sheet =1, cell_rows(6:23))
marg <- read_excel("product_margins.xlsx", sheet =1, cell_rows(6:23))

desc1 <- desc[, -c(2,5)]
marg1 <- marg[, -2]

desc_marg <- merge(desc1, marg1)
names(desc_marg) <- c("Product", "Parent_Co", "Release_Date", "Talk_Time", "Weight",
                      "Screen", "Memory", "Camera", "PPI", "Margin")

head(desc_marg)
```

```
##   Product  Parent_Co Release_Date Talk_Time Weight Screen Memory Camera
## 1 Blade 1 Voltswitch   2011-03-01        20   11.8   4.70     10    5.0
## 2 Blade 2 Voltswitch   2012-07-01        25    8.0   5.00     10    6.7
## 3 Blade 3 Voltswitch   2013-03-01        20    6.1   5.40     12    6.7
## 4 Blade 4 Voltswitch   2013-07-01        25    5.7   5.40     12    9.2
## 5 Blade 5 Voltswitch   2014-03-01        20    5.4   5.60     18    8.0
## 6 Blade 6 Voltswitch   2014-07-01        15    4.0   5.75     16    8.0
##   PPI Margin
## 1 198    0.4
## 2 223    0.6
## 3 265    0.4
## 4 265    0.6
## 5 297    0.4
## 6 297    0.6
```

```
dm_data <- desc_marg %>%
  select(Talk_Time, Weight, PPI, Camera)

prod_preds <- predict(tester_model, dm_data)
desc_marg$pred <- prod_preds
```

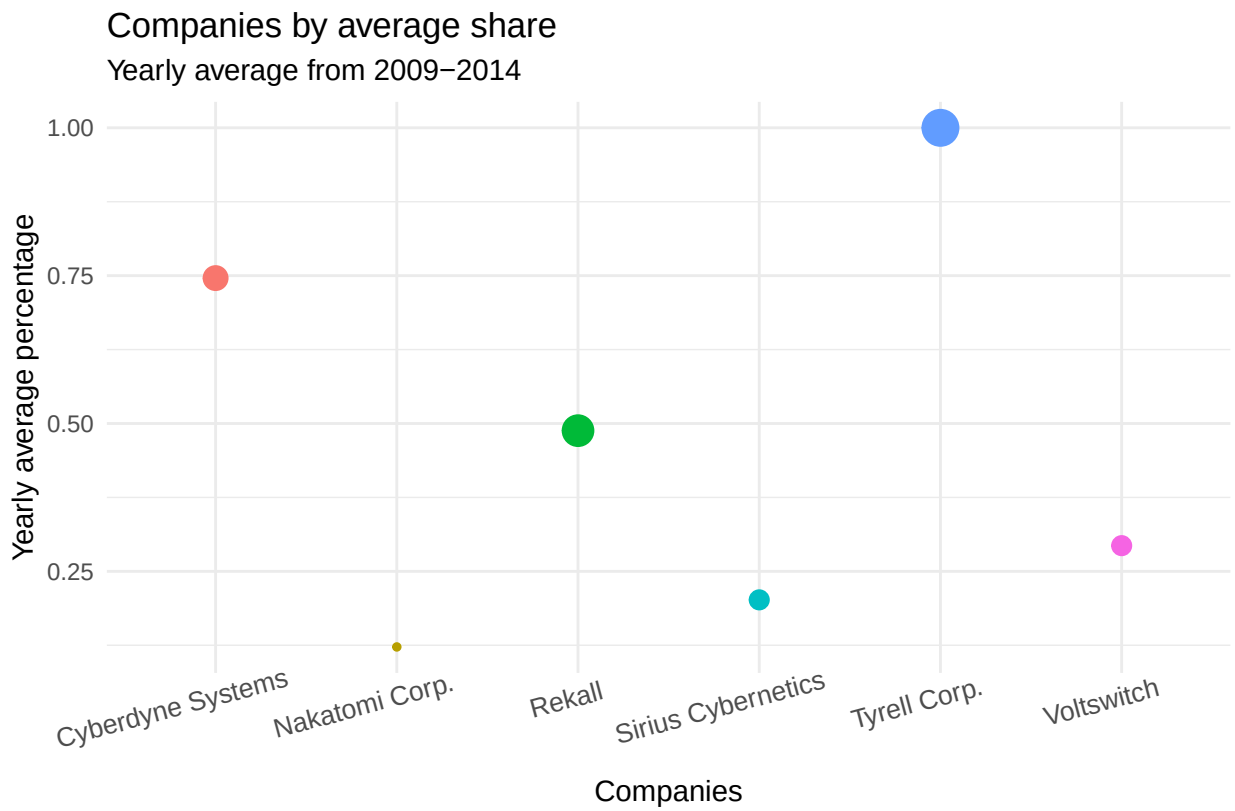
```
desc_marg <- desc_marg %>%
  mutate(profits = pred * Margin) %>%
  mutate(pred_royalty = pred * 0.02) %>%
  mutate(profit_over_license = profits - pred_royalty) %>%
  arrange(profits)

tail(desc_marg, n= 9)
```

```
##     Product      Parent_Co Release_Date Talk_Time Weight Screen Memory
## 9   Blade 6     Voltswitch   2014-07-01        15    4.0   5.75     16
## 10  Blade 2     Voltswitch   2012-07-01        25    8.0   5.00     10
## 11  Blade 4     Voltswitch   2013-07-01        25    5.7   5.40     12
## 12  Mycro 1     Voltswitch   2009-07-01        35    8.0   4.50      8
## 13  Mycro 3     Voltswitch   2011-07-01        35    8.2   5.00     10
## 14  Mycro 2     Voltswitch   2010-07-01        40    8.0   4.20     10
## 15 HAL 1000 Lyte Solutions   2012-05-01        50    9.0   4.50     10
## 16 HAL 2000 Lyte Solutions   2013-05-01        50    9.0   4.50     10
## 17 HAL 3000 Lyte Solutions   2014-05-01        60    9.0   4.50     12
##    Camera PPI Margin     pred  profits pred_royalty profit_over_license
## 9     8.0 297    0.6 224.4371 134.6623     4.488742            130.1735
## 10    6.7 223    0.6 257.2491 154.3494     5.144981            149.2045
## 11    9.2 265    0.6 304.4927 182.6956     6.089854            176.6058
## 12    2.0 174    0.6 321.5335 192.9201     6.430670            186.4894
## 13    5.0 207    0.6 342.6130 205.5678     6.852260            198.7155
## 14    4.2 180    0.6 384.0105 230.4063     7.680211            222.7261
## 15    5.0 200    0.5 482.3491 241.1746     9.646982            231.5276
## 16    5.0 223    0.5 488.0474 244.0237     9.760947            234.2627
## 17    5.0 264    0.5 598.1587 299.0793    11.963174            287.1162
```

Validation through visualization

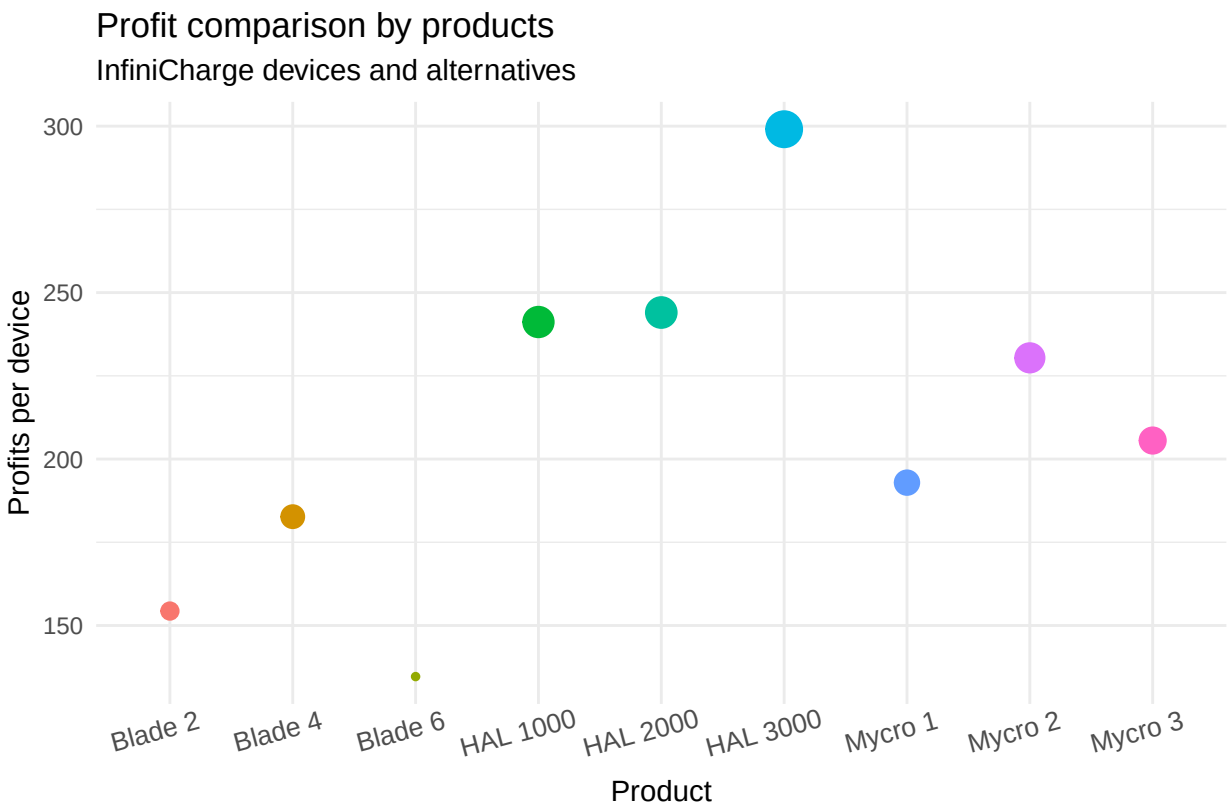
We create some visuals which convey the structure of the data to non-technical stakeholders

```
reg_tot3 %>%
  ggplot(aes(x = company, y = avg_percent, size = royalty_percent)) +
    guides(size = FALSE, colour = FALSE) +
    labs(x = "Companies",
         y = "Yearly average percentage",
         title = "Companies by average share",
         subtitle = "Yearly average from 2009-2014",
         caption = "We see that generally the royalty rate increases as the percentage increases.") +
    geom_point(aes(colour = company)) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 15, hjust = 0.7, size = 10))
```



We see that generally the royalty rate increases as the percentage increases.

```
prof_dev <- desc_marg[9:17,]

pd_viz_df <- prof_dev[, c(1, 12)]

pd_viz_df %>%
  mutate(Product = as.factor(Product)) %>%
  ggplot(aes(x = Product, y = profits)) +
    guides(size = FALSE, colour = FALSE) +
    labs(x = "Product",
         y = "Profits per device",
         title = "Profit comparison by products",
         subtitle = "InfiniCharge devices and alternatives",
         caption = "We see that the profits of using InfiniCharge are greater per device than alternati
    geom_point(aes(colour = Product, size = profits)) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 15, hjust = 0.7, size = 10))
```



We see that the profits of using InfiniCharge are greater per device than alternative