AB Testing Analysis

Matteo Montrucchio

2024-04-08

In the following file, I will evaluate the effectiveness of featuring actions. I will analyze a dataset characterized by weekly sales and prices of orange juices across different stores for three brands. To evaluate the effectiveness of featuring actions, I compare the average sales by feat category.

```r
library(readr)
oj.data <- read_csv("/Users/matteomontrucchio/Desktop/oj_data.csv", show_col_types = FALSE)
```

```r
# show the first rows of the dataset
head(oj.data)
```

```
## # A tibble: 6 x 4
##   sales price brand      feat
##   <dbl> <dbl> <chr>     <dbl>
## 1 8256.  3.87 tropicana     0
## 2 6144.  3.87 tropicana     0
## 3 3840.  3.87 tropicana     0
## 4 8000.  3.87 tropicana     0
## 5 8896.  3.87 tropicana     0
## 6 7168.  3.87 tropicana     0
```

```r
# summarize the information contained in the dataset
summary(oj.data)
```

```
##      sales             price           brand                feat
##  Min.   :    64   Min.   :0.520   Length:28947       Min.   :0.0000
##  1st Qu.:  4864   1st Qu.:1.790   Class :character   1st Qu.:0.0000
##  Median :  8384   Median :2.170   Mode  :character   Median :0.0000
##  Mean   : 17312   Mean   :2.282                      Mean   :0.2373
##  3rd Qu.: 17408   3rd Qu.:2.730                      3rd Qu.:0.0000
##  Max.   :716416   Max.   :3.870                      Max.   :1.0000
```

```r
# generate the plot comparing average sales by feat category and their SD
oj.data %>%
  mutate(feat= as.factor(feat)) %>%     # from the summary(), feat is stored as numeric variable
  group_by(feat) %>%
  summarise(Average = mean(sales),
            SD = sd(sales)) %>%
  ggplot(aes(feat,Average)) +
  geom_point(size=5) +
  geom_errorbar(aes(ymin=Average-SD, ymax=Average+SD), width=.2)
```
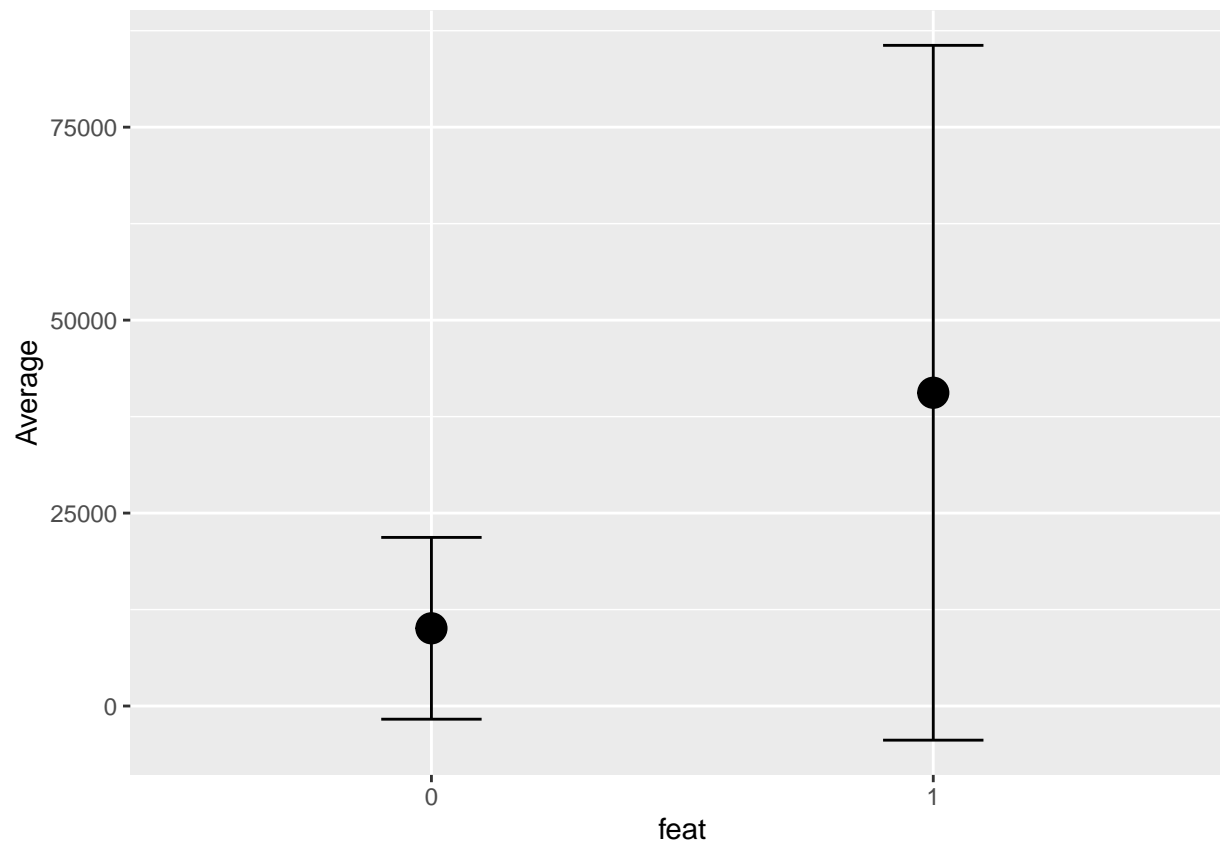
```
t.table <- oj.data %>%
  mutate(feat= as.factor(feat)) %>%
  group_by(feat) %>%
  summarise(Average = mean(sales),
            Var = var(sales),
            N = n())
t.table
```

```
## # A tibble: 2 x 4
##   feat  Average           Var     N
##   <fct>   <dbl>         <dbl> <int>
## 1 0       10071.   138704271. 22079
## 2 1       40590.  2026106369.  6868
```

```
t.test(oj.data[oj.data$feat==1, 1], oj.data[oj.data$feat==0, 1], var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.data[oj.data$feat == 1, 1] and oj.data[oj.data$feat == 0, 1]
## t = 55.601, df = 7161.6, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  29443.30 31595.31
## sample estimates:
```

```
## mean of x mean of y
##   40590.47  10071.17
```

The data provided shows how average sales are higher when the product is featured.

On average, 40590 quantities are sold when product is featured against an average of 10071 quantities sold when not featured, resulting in an average difference of 30519 quantities (with a 95% probability, a store will sell between 29443.16 and 31594.84 more quantities).