MSBA 6440 Homework 1

Team 11

Jiaru Cai, Yuwen Yan, Yuxuan Wang, Nakul Agrawal, Pavan Bapesh Vinnakota

Feb 11, 2020

Table of Contents

Introduction and Overview

Star digital is a multichannel video service provider with over US$100 million in annual advertising spends. Since 2012/13, Star Digital started gradually increasing the share of online advertising spend but found that measuring display ad effectiveness is difficult. To measure the causal effect of display advertising on sales conversion, Star Digital designed a controlled experiment to measure what users would have done if they had not seen the campaign ads. This report analyzes the experiment, validates the assumption and assesses results regarding the success level of the experiment.

Experimental Design

Star Digital is experimenting on whether their ads are effective in influencing customers to purchase packages. They conducted an A/B testing on two groups of customers. Under controlled experiment, the customers are randomly assigned to the 'test' group in which they can see the advertisement from Star Digital's advertising campaign, and 'control' group in which they only see the charity advertisement in place of the Star Digital advertisement.

Star Digital have over 25,000 customers. They considered some factors and determined that in its experiment, 10 percent of the consumers would be placed in the control group, and the remaining 90 percent would be placed in the test group. The control group was designed to be smaller considering to minimize the opportunity cost for the experiment.

Conditions:

1. Randomly assigned users to control and treatment group.

2.They carefully decided on the proportion of consumers to be assigned to the control group, so as to get a statistically valid comparison at minimum cost.

3.The outcome is reliable.

Randomization Test

Before analyzing the experiment's result, we would first conduct randomization tests to check the purchase behavior of control and treatment groups were not influenced by other factors. Since the advertiser cannot choose website 1 through 5 but it can choose from

website 1 through 5 and website 6, we treat website 1 to 5 as the same and conduct t tests on “sum1to5” and “imp_6” against “test”.

```
# Load the packages
library(dplyr)

library(ggplot2)
library(MatchIt)
library(rcompanion)
library(pwr)

# Read the dataframe
data <- read.csv('starDigital.csv')

# Randomization Test
t.test(sum1to5 ~ test, data = data)

##
##  Welch Two Sample t-test
##
## data:  sum1to5 by test
## t = -0.071371, df = 3268.6, p-value = 0.9431
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8402427  0.7812196
## sample estimates:
## mean in group 0 mean in group 1
##        6.065512        6.095024
```

Interpretation - The null hypothesis for the t-test above is there is no difference between average impression on website 1 to 5 for treatment and control groups. The alternative hypothesis is there is a difference between average impression on website 1 to 5 for treatment and control group. The p-value of the test is 0.9431 which gives a really weak evidence to reject the null hypothesis.

```
t.test(imp_6 ~ test, data = data)

##
##  Welch Two Sample t-test
##
## data:  imp_6 by test
## t = 0.43156, df = 2898.4, p-value = 0.6661
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.3176712  0.4969729
## sample estimates:
## mean in group 0 mean in group 1
##        1.863705        1.774054
```

Interpretation - Similar to the first t-test, the null hypothesis is there is no difference between average impression on website 6 for treatment and control group, and the alternative hypothesis is there is a difference between average impression on website 6 for treatment and control groups. The p-value is also large enough so that we cannot reject the null hypothesis.

Conclusion - Both treatment and control groups got similar exposure to the advertisement in all six websites.

Question1. Is online advertising effective for Star Digital?

After rigorous experimentation, we used logistic regression to test the advertising effect between control and treatment group, that is whether the Star Digital advertisement would increase customers' purchases.

```
m1 <- glm(purchase ~ test, data =  data, family = "binomial")
summary(m1)
```

```
##
## Call:
## glm(formula = purchase ~ test, family = "binomial", data = data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.186  -1.186   1.169   1.169   1.202
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.05724    0.03882  -1.474   0.1404
## test         0.07676    0.04104   1.871   0.0614 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 35073  on 25301  degrees of freedom
## AIC: 35077
##
## Number of Fisher Scoring iterations: 3
```

Interpretation - The p-value shown in the result is 0.06, which is pretty small so we would reject our null hypothesis and accept that the online advertising is effective for Star Digital. When significance level is less than or equal to 10%, we have at least 90% of confidence to believe that on the average, every unit increase of online advertising will increase chance of purchase by 7.68%.

Question 2. Frequency effect of advertising on purchase?

Here we separate all the websites and take each one's frequency into consideration by doing a logistic regression to calculate them.

```
data$freq <- data$imp_1 + data$imp_2 + data$imp_3 + data$imp_4 + data$imp_5 +
 data$imp_6
m2 <- glm(purchase ~ freq + test*freq, data = data, family = "binomial")
summary(m2)
```

```
##
## Call:
## glm(formula = purchase ~ freq + test * freq, family = "binomial",
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.9145  -1.1266   0.1299   1.2156   1.2433
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.169577   0.042895  -3.953 7.71e-05 ***
## freq         0.015889   0.002876   5.524 3.32e-08 ***
## test        -0.013903   0.045613  -0.305    0.761
## freq:test    0.015466   0.003207   4.823 1.42e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 34190  on 25299  degrees of freedom
## AIC: 34198
##
## Number of Fisher Scoring iterations: 5
```

From the model result above, first, we find that every unit increase of frequency in the website advertisement impression will increase chance of buying by 1.59%, while holding the other factors fixed. Second, there is an extra impact towards people especially when they are exposed to the Star Digital advertisement. On the average, every unit increase of the Star Digital Advertisement will increase the chance of purchase by 1.55% compared with people who didn't see ads from Star Digital.

Question 3. Which sites should Star Digital advertise on?

For the last part, we use regression model to investigate whether impressions from different website would cause different effect on purchase. While the advertiser cannot control which of Sites 1 through 5 it can advertise on, it does have the ability to specify if the advertising should appear on Site 6 or Sites 1 through 5, or both these options. Therefore, we regress "purchase" on "test", "sum1to5", "imp_6" and their interactions.

```
m3 <- glm(purchase ~ test + sum1to5 + imp_6 + test*sum1to5 + test*imp_6, data
 = data, family = "binomial")
summary(m3)
```

```
##
## Call:
## glm(formula = purchase ~ test + sum1to5 + imp_6 + test * sum1to5 +
##     test * imp_6, family = "binomial", data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.1280  -1.1195   0.1185   1.2217   1.2472
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  -0.166556   0.042533  -3.916 9.01e-05 ***
## test         -0.006087   0.045314  -0.134 0.893139
## sum1to5       0.019452   0.003443   5.650 1.61e-08 ***
## imp_6         0.003978   0.004294   0.927 0.354179
## test:sum1to5  0.014617   0.003794   3.852 0.000117 ***
## test:imp_6    0.013483   0.005405   2.494 0.012616 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 34166  on 25297  degrees of freedom
## AIC: 34178
##
## Number of Fisher Scoring iterations: 5
```

Interpretation - The coefficients of the interaction terms (test * sum1to5, test * imp_6) can imply the effect of Star Digital's impression at different websites on purchase. All those coefficient are significant, which indicates that one more impression at Website 1 through 5 would increase the probability of purchase by 1.5% compared to people who didn't see the ads from Star Digital. And one more impression at Website 6 would increase the probability of purchase by 1.3% compared to people who didn't see the ads from Star Digital. To decide which website should Star Digital advertise on, we need to take the ROI into consideration.

```
# ROI for one ad at Web 1 to 5
ROI_1to5 <- (1200 * .015 - 25 / 1000) / (25 / 1000)
ROI_6 <- (1200 * .013 - 20 / 1000) / (20 / 1000)
ROI_1to5
```

```
## [1] 719
```

```
ROI_6
```

```
## [1] 779
```

If we only consider the return on each advertisement, website 1 through 5 would be a better choice. However, we'd better choose ROI as the decisive criterion, which can take the

return and cost into consideration. Comparing the value of ROI, Star Digital should advertise on website 6.

Summary

- The experiment is designed properly with matching groups in both control and treatment.
- Under the 10% significance level, we have 90% of confidence to think that the Star Digital advertising is effective.
- Frequency indeed has an impact on both of normal advertisement and the Star Digital advertisement.
- For the advertisement investment, if the Star Digital simply thinks about the revenue, website 1-5 is the best choice. However, if ROI is the first priority, then the company should choose the website 6.