6440 - Homework 1 - Star Digital

Prasanna Rajendran, Divya Upadhyay, Keming Hu, Wade Wimer, Meng (Julius) Yang

February 12, 2019

Contents

Introduction and overview:

Star Digital is a multi-channel video service provider. They spent over US$100 million in annual advertising spends and are looking to increase online advertising spend to futher increase sales. Even though the Internet provides a promising platform to measure ad effectiveness, doing so for display ads had proved difficult in attempts thus far. The two common approaches to measuring display ad effectiveness tended to either under value or over value conversion rates. To truly measure the effect on sales conversion, Star Digital designed a controlled experiment to measure the effect of display ads for one of its advertising campaigns. This paper describes an analysis of the experiment, validating the assumptions and choices that were made by focusing on the experiment design and answering some of the important questions about the success level of the results.

Executive Summary:

Star Digital carefully designed an experiment to test the effect of various online advertising channels on sales using a sample set of over 25,000 online customers. Most of the experimental design suggests that the test was carried out with no apparent reason that a causal conclusions could not be drawn. The one primary area of concern is that our analysis here was drawn from a small sample of the original experimental data set that could lead to selection bias. Upon testing for similarity between test and control groups, the data indicates that the number of impressions per customer between the groups are statistically similar, suggesting that we can be confident that two groups are identical. Upon examination of the amount of purchases in the control group versus the test group, our findings suggested that the group that received the Star Digital online ads, had a higher incidence of purchases than the group receiving the charity ads. Additionally, an examination of the number of individuals in each of the control and test group indicated that not enough data points were gathered to put forth a confident conclusion. Despite this, we went on to present our findings with a suggestion of caution. Upon examining the relationship between the number of impressions on the instances of purchase, we found that a 1% increase in number of impressions is having a positive impact on making a customer 0.13% more likely to purchase StarDigital's subscription package. Overall, the benefit gained and costs spent on using websites 1 to 5 suggests in a greater benefit than using website 6. Thus, we recommend that Star Digital apply their online advertizing budget to websites 1 to 5 in order to maximize the chance of conversion based on advertisements.

Experimental Design:

Star Digital is experimenting on whether their ads are effective in influencing customers to purchase packages and increase sales. A test was conducted by creating an A/B experiment on two sets of customers. The unit of analysis is a customer. The treatment group was a 90% portion of the over 25,000 customers. These customers were shown online ads about Star Digitial, while the remaining 10% of customers was designated the control group and were shown ads about charity. The control group was designed to be smaller in order to minimize the opportunity cost sacrificed for the experiment while still maintaining a group large enough to draw statistical conclusions from.

The experiment carried out looks reasonable while looking at some of the factors that Star Digital considered in setting up.

- Star Digitial randomly assigned online users to test and control groups
- The outcome metric can be measured reliably
- The unequal split between the test and control groups was intentionally planned.

Despite several aspects of the experiment being well planed, We feel the experimental design could be improved by the following factors:

- There is no mention of the level of foresight customers had about the experiment. Knowing about the experiment could lead to change in behavior

Threats to causal inference:

We examined the data and experiment to findi confidence that there is no threat to causal inference and if there is any, we would like to call it out:

- The data provided is a sample of the original experiment data set. This is not explained in our instructions and nor are the details of the sampling method of this data set provided. Selection bias could be a factor in ths work. For example,the sample could be favoring towards purchases in either of the groups.
- There is no observable simulataneity effect going into the experiment. In this case, only impressions affect purchases and there is no reason specified for us to believe purchases would lead to more impressions
- Since this an experimental data, we believe that there is no measurement error. Still we feel its better if the measurements that are recorded are checked for its authenticity once after the experiment if not already completed.

Descriptive Statistics:

We begin with understanding the star digital data that comprises of customers and impressions from the six websites. In addition to that, data provides purchase(0 or 1) and test-control(0-control, 1-test) for each customer.

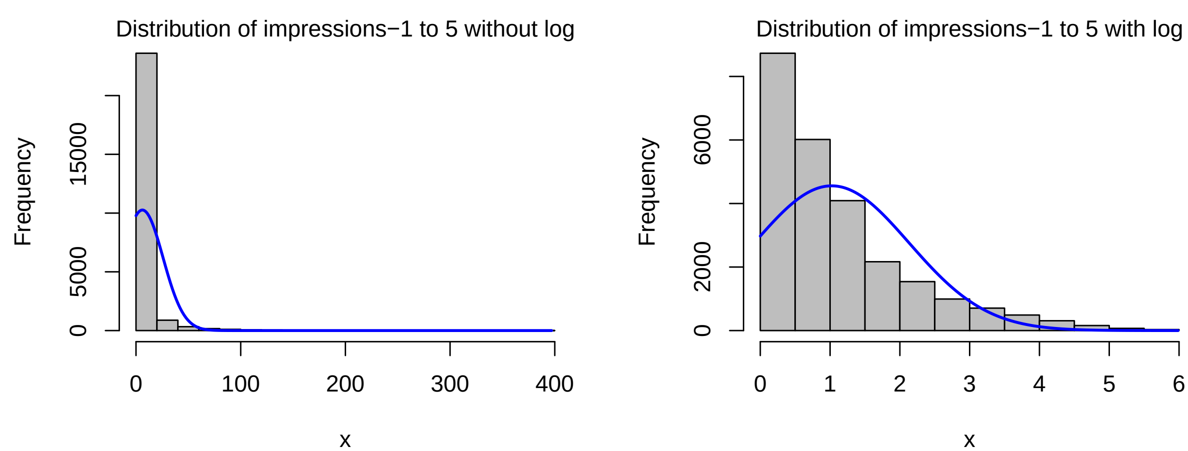
```r
# Importing necessary packages
suppressPackageStartupMessages({
  library(TSA);  library(forecast);  library(ggplot2);  library(dplyr);
  library(stargazer); library(ggplot2);  library(data.table);  library(tableone);
  library(lattice); library(MESS); library(pwr);  library(rcompanion);
  library(scales);  library(plm)
})
```

Loading the data in R and and observing the summary statistics for all columns, we observed that the impressions across websites might be highly skewed. To under this better we look at individual distributions.

```r
# reading the dataframe
star <- as.data.frame(
  fread('/01 Drive/UMinn/Spring/Data-Driven Exp/Homeworks/HW1/starDigital.csv'))
```

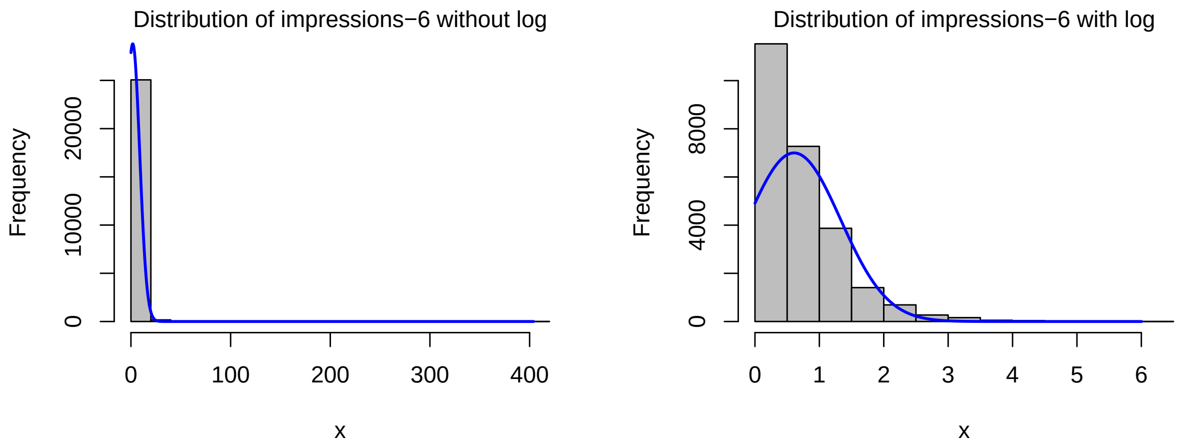
Plotting a histogram for impressions from website 1-5 shows a strong negative skew. Taking log transform and plotting it again, gives a better distribution

```r
# Creating histogram with distribution of impressions from website 1 -5
# with and without log transformation
par(mfrow=c(1,2))
plotNormalHistogram(star$sum1to5)
mtext("Distribution of impressions-1 to 5 without log")
plotNormalHistogram(log(star$sum1to5 + 1))
mtext("Distribution of impressions-1 to 5 with log")
```



Similarly, plotting a histogram for impressions from website 6 also shows a strong negative skew. So we again take a log transform for this

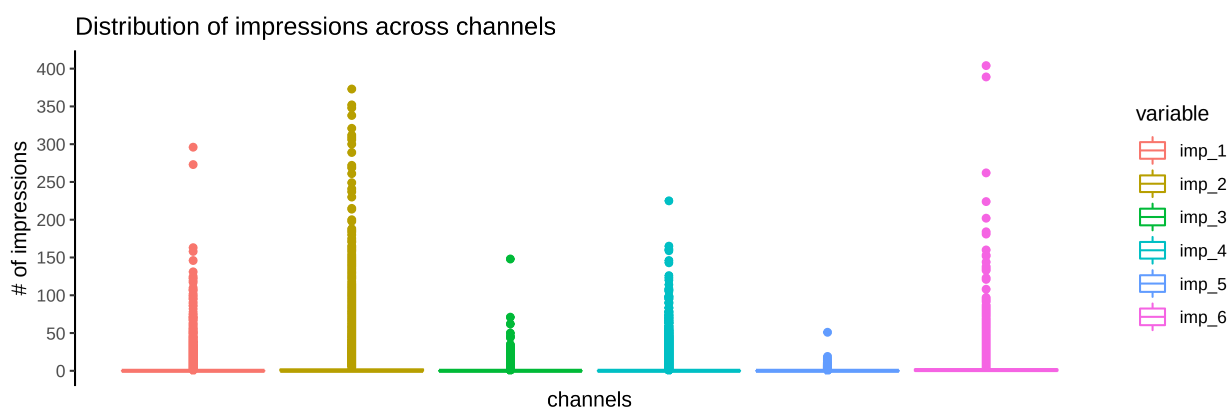
```r
# Creating histogram with distribution of impressions from website 6
# with and without log transformation
par(mfrow=c(1,2))
plotNormalHistogram(star$imp_6)
mtext("Distribution of impressions-6 without log")
plotNormalHistogram(log(star$imp_6+1))
mtext("Distribution of impressions-6 with log")
```



From websites 1 to 5, we would like to see the distribution of the individual websites and see if there is any website that is not different, so plot a box plot for all the websites individually.

```
# Plotting a box plot for each websites

long = melt(star,id.vars = "id",measure.vars =
              c("imp_1","imp_2","imp_3","imp_4","imp_5","imp_6"))
long$variable <- as.character(long$variable)
ggplot(aes(y=value, colour=variable), data=long) +
  geom_boxplot() +
  scale_y_continuous(breaks = c(seq(0,450,50))) +
  scale_x_continuous(name="channels",breaks=seq(1,6,1)) +
  ylab("# of impressions") +
  ggtitle("Distribution of impressions across channels") +
  theme_classic()
```



We can see that the 1st quartile, median and the 3rd quartile all are around 0 and only some observations have values above them

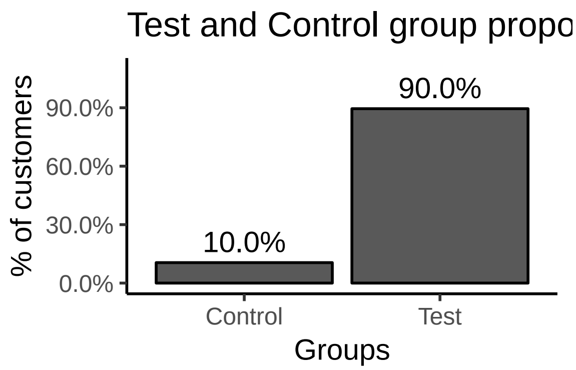
Its said that the test and control split is a 90:10 split. We would like to test this statement by plotting the distribution

```
# Calculating proportion between test and control group
```

```r
star %>%
  count(test) %>%
  mutate(perc = n / nrow(star)) -> test_stat

# Plotting the proportion between control and test groups

ggplot(data=test_stat, aes(x= factor(test), y=perc)) +
  geom_bar(colour="black", stat="identity") +
  geom_text(aes(label= scales::percent(round(perc,2)), vjust=-.5)) +
  ylab("% of customers") +  xlab("Groups") +
  scale_y_continuous(labels=percent_format(), limits = c(0,1.1)) +
  scale_x_discrete(labels=c('Control', 'Test')) +
  ggtitle("Test and Control group proportion") + theme_classic()
```



The size of test and control is as follows:

```r
# getting the count of control and test groups
table(star$test)
```

```
##
##     0     1
##  2656 22647
```

We have 2656 customers in control and 22647 customers in test group

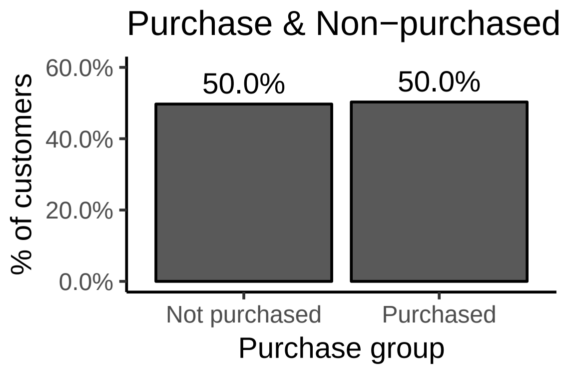
We also would like to validate whether the purchased proportion of 50:50 is true or not by plotting the proportion

```r
# Calculating proportions of purchases

star %>%
  count(purchase) %>%
  mutate(perc = n / nrow(star)) -> pur_stat

# Plotting the proportion between Purchased and Non-purchased groups

ggplot(data=pur_stat, aes(x= factor(purchase), y=perc)) +
  geom_bar(colour="black", stat="identity") +
  geom_text(aes(label= scales::percent(round(perc,2)), vjust=-.5)) +
  ylab("% of customers") +  xlab("Purchase group") +
  scale_y_continuous(labels=percent_format(), limits = c(0, 0.6)) +
  scale_x_discrete(labels=c('Not purchased', 'Purchased')) +
  ggtitle("Purchase & Non-purchased proportion") + theme_classic()
```



The statement on purchased or not proportion to be 50: 50 is validated by from the bar graph shown above.

Randomization Test

We would first check whether subjects in two groups are not influenced by other factors. To conduct the randomization check, we performed t.test on all the variables individually against the test variable.

```
#Using t-test to do randomization check
t.test(sum1to5 ~ test,data=star)
```

```
##
##  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
```

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

```
##
##  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
```

Among the two aforementioned tests,the first test checks if there is any bias between test and control groups with respect to impressions from website 1-5. Being a statistical test, the null hypothesis and alternate hypothesis are as follows: Null hypothesis: There is no difference between the (true) averages of impressions between test and control group Alternate hypothesis: There is a difference between the (true) averages of impressions between test and control group

The test results show a significance of 0.9431, which implies a very weak evidence against the null hypothesis, hence we fail to reject the null hypothesis. This interpretation can also be observed by noticing the mean impressions in the two groups are extremely similar (mean in group $0: 6.065512$, mean in group $1: 6.095024$). We observe similar output from the second test which checks if there is any bias between test and control groups with respect to impressions from website 6.

Concluding from both the t.test and histograms, we could find that subjects in both groups got a highly similar exposure to all 6 sites.

Significance test:

The preliminary intention of this experiment is to test whether the money Star Digital spent on online ads have actually translated into purchases. To check whether Star Digital make a difference, we have to test whether test results in two groups are significantly different. Here, we will use t-test to testify whether average number of purchases of test group is significantly higher than that of control group.

```r
t.test(purchase~test,data = star)
```

```
##
##  Welch Two Sample t-test
##
## data:  purchase by test
## t = -1.8713, df = 3309.2, p-value = 0.06139
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.039289257  0.000916332
## sample estimates:
## mean in group 0 mean in group 1
##       0.4856928       0.5048792
```

Comparing the mean purchase proportion in test and control group, we see from the output of t.test above that the control group has a mean purchase proportion of 0.48 while the test group has a mean purchase proportion of 0.50. With an approximate difference in mean being 0.019, our next statistical test should check whether the available sample size is sufficient to reliably detect the difference in mean of 0.019 between the test and control groups.

But we would also like to validate this significance by making sure we have enough sample size available in the given data set that is sufficient to detect change in purchase with a given degree of confidence. So we conduct significance test between test and purchase to find the mean difference

To check the sufficiency of data points, we use the power_t_test function and specify the arguments required. From the descriptive analysis of data set, we know that the control set is 1/9th in size as compared to the test set, hence ratio argument is specified as 9.

```r
# Calcuting the sample needed to satisfy the above delta
power_t_test(n=NULL,
             type=c("two.sample"),
             alternative="two.sided",
             ratio = 9,
             delta=0.019,
             power=0.8,
             sig.level = 0.1)
```

```
##
##      Two-sample t test power calculation with unequal sizes
##
##               n = 19030.21, 171271.85
##           delta = 0.019
```

```
##              sd = 1, 1
##       sig.level = 0.1
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is vector of number in each group
```

The test results state that we require at least 19030 samples in control set and 171271 samples in test set in order to reliably detect a difference in mean of 0.019 with 90% confidence. Therefore, our experiment appears to be heavily under powered to detect the effect management is looking for. Thus, we would caution the management about reading too much into results from this experiment and re-run the experiment with a larger sample data set.

Relationship between impressions and purchase

Now that we have established that there is a difference in relationship between the test group and the control group, we would like to see what kind of relationship they hold and to what level. In other words, we would like to see how the impressions relate with the rate of purchase.

We use regression models to find out whether the change in number of impressions would results in changes in purchase. And because the distribution of impression amount is skewed heavily, we regressed purchase on log of total impressions to remove potential effect of skewness.

```r
#first, creat a new column named total to calculate the total impression number across all sites
star$total <- star$imp_6 + star$sum1to5

#second, using total number of impressions to check effect from the frequency of ads
total <- plm(purchase ~ total, data = star, index = 'test',
             effect = 'individual', model = 'within')
#summary(total)

lg_total <- plm(purchase ~ log(total + 1), data = star,
                index = 'test', effect = 'individual', model = 'within')
#summary(log_total)

stargazer(total, lg_total,
          title='Regress purchase on total and log(total)',
          type = 'text',column.labels = c('total','log(total)'))
```

```
##
## Regress purchase on total and log(total)
## ========================================================
##                                 Dependent variable:
##                             ----------------------------
##                                       purchase
##                                 total       log(total)
##                                  (1)           (2)
## --------------------------------------------------------
## total                          0.004***
##                                (0.0001)
##
## log(total + 1)                               0.129***
##                                               (0.003)
##
## --------------------------------------------------------
```

```
## Observations                      25,303        25,303
## R2                                0.023          0.060
## Adjusted R2                       0.023          0.060
## F Statistic (df = 1; 25300)   590.911***   1,615.799***
## ===========================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

Interpretation From the two regression models, we found out that regression with log total captures higher variance, as the R squared is higher, which means log of total number of impressions performs better in predicting or explaining purchases. And from the result we got, we could know that 1% increase in number of impressions would expect to make a customer 0.13% more likely to purchase StarDigital's product. And it subsequently implies that frequency of ads impressions would have a positive influence on customer purchasing behavior.

We took guidance on interpretation of log variable from the link specified below:

Link for interpretation

Choosing between website 1 to 5 and website 6

Now that we have established the effect of impressions on purchase, we would like to dig deeper and find the impacts on advertising in different websites and potentially find which website is better to advertise by including the cost.

We used regression models to measure the effect of number of impressions from different websites on changing purchase behavior. Based on the cost of of advertising at sites, we can finally calculate the cost of changing the probability of purchase, and we utilize our calculation to provide the recommendation for the website Star Digital should invest on

```r
# Building a model for impressions from website 1-5 as independent
within_reg <- plm(purchase ~ sum1to5,
              data = star, index = c('test'), model = 'within',
              effect = 'individual')

# Building a model for impressions from website 6 as independent
within_reg1 <- plm(purchase ~ imp_6,
              data = star, index = c('test'), model = 'within',
              effect = 'individual')

stargazer(within_reg,within_reg1,type="text")
```

```
##
## ===========================================================
##                                    Dependent variable:
##                                ----------------------------
##                                          purchase
##                                    (1)            (2)
## -----------------------------------------------------------
## sum1to5                          0.004***
##                                 (0.0002)
##
## imp_6                                           0.003***
##                                                 (0.0004)
##
## -----------------------------------------------------------
## Observations                      25,303        25,303
```

```
## R2                             0.022          0.002
## Adjusted R2                    0.022          0.002
## F Statistic (df = 1; 25300)   582.048***     48.106***
## ==========================================================
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

The fixed effect regressions show that 1 unit increasing of impression from website 1 to 5 will result in 0.004 probability increase in purchase, and 1 unit increasing of impression from website 6 will result in 0.003 probability increase in purchase, which both of them are significant with respective to the p value.

```
#calculate the cost per unit impression on website 1 to 5 and website 6:
cost_sum1to5 <- 25/1000
cost_imp_6 <- 20/1000

#calculate the cost for 1 possibility increasing of purchase:
cost1 <- cost_sum1to5/0.004 # cost for website 1 to 5
cost2 <- cost_imp_6/0.003 # cost for website 6

cost1
```

```
## [1] 6.25
```

```
cost2
```

```
## [1] 6.666667
```

Regarding the cost of advertising on different website for one thousand impressions, the cost per unit impression is $0.025 for website 1 to 5, and on the other hand, the cost per unit impression for website 6 is $0.02. Then we calculate the cost for 1 unit increase in purchase, which is: 6.25 for website 1 to 5, and 6.67 for website 6.

The cost for website 1 to 5 for 1 percent probability increasing of purchase is cheaper than website 6, so we recommend that Star Digital invest money on increase the number of impressions on website 1 to 5 instead of website 6.

Summary of Recommendations:

Based on the evaluation the experimental design and statistical analysis of data, we can conclude following inferences from the results of this experiment:

- Caution regarding sample size and power of experiment: In order to reliably detect the difference in mean of purchases between test and control sets, the given sample size is not sufficient. Even though the statistical tests show that impressions have positive relationship with purchase probability, we cannot be confident of the experiment results due to under powered sample size. We recommend sample size of at least 19,030 in the control set (i.e. smaller among test and control), to reliably detect the difference in purchase behavior in test set compared to control set.
- Effect of impressions on purchase likelihood: From the statistical analysis of the data, we can conclude that customers in the test set, who see Star Digital ads instead of charity ads have a higher likelihood of making a purchase. It was also concluded that higher frequency of impression will have an incremental effect on purchase likelihood.
- Profitability across websites: Website 1 to 5 seem to be more profitable as compared to website 6 since 1 unit increase in purchases would incur a cost of $6.25 for websites 1 to 5 while, website 6 would incur $6.67.