# Online Campaign Evaluation

*Group: Fu-Yi Pao (Maggie), Robert Zhang*

## 1. Randomized Check

### a. Verify the randomization by user country location

### i. Use the difference-in-means estimator to compare the proportions of UK and US consumers in Treatment versus Control groups.

The difference in means for the "IsUK" variable is 0.017%, with a standard deviation of 0.044%. This indicates that the treatment group had a slightly higher proportion of UK consumers than the control group. While, the difference in means for the "IsUS" variable is -0.019%, with a standard deviation of 0.05%, indicating that the control group had a slightly higher proportion of US consumers than the treatment group.

### ii. The validity of the randomization in terms of user location

The 95% confidence interval for the difference in means of 'IsUK' is (-0.000699, 0.001034), which includes zero. Similarly, the 95% confidence interval for the difference in means of 'IsUS' is (-0.001172, 0.000788), which also includes zero. Both results suggest that the difference is not statistically significant, indicating that we failed to reject the null hypothesis that there is no difference in the proportions of UK and US consumers between the treatment and control groups.

```
        Welch Two Sample t-test

data:  df$IsUK[df$treatment == 1] and df$IsUK[df$treatment == 0]
t = 0.4, df = 3e+06, p-value = 0.7
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.000699  0.001034
sample estimates:
mean of x mean of y
     0.25      0.25
```

```
        Welch Two Sample t-test

data:  df$IsUS[df$treatment == 1] and df$IsUS[df$treatment == 0]
t = -0.4, df = 3e+06, p-value = 0.7
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.001172  0.000788
sample estimates:
mean of x mean of y
      0.6       0.6
```

### b. Verify the randomization by past sales

### i. Use the difference-in-means estimator to compare the average sales in the 2 weeks before the experiment in the Treatment versus Control groups.

The result shows that the difference in means for "past_sales" is $0.004, with a standard deviation of $0.005, suggesting that the treatment group had more sales in the 2 weeks before the experiment compared to the control. The 95% confidence interval is (0.013, -0.006).

### ii. The validity of the randomization in terms of past sales

The 95% confidence interval for the difference in means of the past sales variable is (-0.0061 ,0.0133), we can see that it includes zero. Therefore,

```
        Welch Two Sample t-test

data:  df$past_sales[df$treatment == 1] and df$past_sales[df$treatment == 0]
t = 0.7, df = 3e+06, p-value = 0.5
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0061  0.0133
sample estimates:
mean of x mean of y
    0.516     0.512
```

we failed to reject the null hypothesis that there is no difference in mean past sales between the treatment and control groups at the 5% level.

## 2. Ad effect estimation on per-user sales without ghost ads?

Under the assumption of no ghost ads effect, the ITT estimate of the ad effect on per-user sales is $0.373, with a standard deviation of $0.008 and ITT Lift is 42.445%, which means that the ad had a positive impact. The 95% confidence interval for the ITT estimate is (0.389, 0.356), which means that it is statistically significant.

## 3. Experiment with ghost ads

### a. Verify the equivalence of Treatment exposed and Control exposed users by user location

The difference in mean for UK exposed users is 0.368% with standard deviation of 0.058%, and for US exposed users, it is 0.771% with standard deviation of 0.066%. Both indicate that users who saw the ads were slightly more likely to be from the UK and US compared to those in the control group. The t-test results also provide the statistical significance of these differences, with the 95% confidence interval for "IsUK" is between (0.00254, 0.00482) and for "IsUS" is between (0.00642,0.00900), suggesting that the difference in means between the treatment and control groups is statistically significant in both the UK and the US.

```
        Welch Two Sample t-test

data:  df_exposed$IsUK[df_exposed$treatment == 1] and df_exposed$IsUK[df_exposed$treatment ==
0]
t = 6, df = 2e+06, p-value = 2e-10
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.00254 0.00482
sample estimates:
mean of x mean of y
    0.250     0.247


        Welch Two Sample t-test

data:  df_exposed$IsUS[df_exposed$treatment == 1] and df_exposed$IsUS[df_exposed$treatment ==
0]
t = 12, df = 2e+06, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.00642 0.00900
sample estimates:
mean of x mean of y
    0.600     0.592
```

The results for all users(no ghost ad) indicate that there is no statistically significant difference between the treatment and control groups in the proportion of UK and US consumers. However, the results for exposed users show a statistically significant difference in the proportion of UK and US consumers.

### b. Verify the equivalence of Treatment exposed and Control exposed users by past sales

The result suggests that there is no significant difference between the past sales for exposed users in the treatment and control group. The difference in means is -0.001$ with standard deviation of 0.007$, which indicates that the average past sales of the treatment exposed group is slightly lower than that of the control. The 95% confidence interval of (0.012, -0.013) indicates that we failed to reject the null hypothesis that there is no significant difference in the past sales of treatment exposed and control exposed users.

Regardless of whether users were exposed to ads or not, the finding suggests that there is no statistically significant difference in past sales between the treatment and control groups. This result is consistent across

```
        Welch Two Sample t-test

data:  df_exposed$past_sales[df_exposed$treatment == 1] and
df_exposed$past_sales[df_exposed$treatment == 0]
t = -0.1, df = 2e+06, p-value = 0.9
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0135  0.0121
sample estimates:
mean of x mean of y
    0.517     0.517
```

both the full user population and the subset of users who saw ads.

**4. Ad effect estimate changes with ghost ads? Compute the Treatment on Treated (TOT) estimate for users who saw ads.**

The sales effect point estimate for the TOT is 0.656$ and the lift is 73.642%, which means that users who were targeted with ads had sales that were 0.656$ higher(with standard error is 0.0118) than users who were not targeted with ads. The 95% confidence interval is (0.633 , 0.679), suggesting that the effect in sales associated with being exposed to ads is statistically significant.

**5. The effect of the ad campaign on total sales (i.e. gross revenue)**

**a. Compute the effect using the ITT estimate.**

```
        Welch Two Sample t-test

data:  df$sales[df$treatment == 1] and df$sales[df$treatment == 0]
t = 45, df = 4e+06, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.356 0.389
sample estimates:
mean of x mean of y
    1.250     0.878

ITT of incremental gross revenue is: 894339
ITT Lift on total sales is: 1.14
S.E. of ITT is: 0.008
```

The point estimate of ITT on total incremental gross revenue is $894,339 (with standard deviation of 0.008), representing the extra gross revenue the treatment group generated compared to the control group. The lift of ITT on gross revenue is 1.138%. The 95% CI is (0.356, 0.389), meaning the lift of ITT on total gross revenue is statistically significant.

**b. Compute the effect using the TOT estimate.**

```
        Welch Two Sample t-test

data:  df$sales[df$treatment == 1 & df$saw_ads == 1] and df$sales[df$treatment == 0 &
df$saw_ads == 1]
t = 56, df = 2e+06, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.633 0.679
sample estimates:
mean of x mean of y
    1.547     0.891

TOT of incremental gross revenue is: 872379
TOT Lift is: 1.428%
SD of TOT is: 0.012
```

After conditioning on both seeing ads, the treatment group generated $872,379 (with standard deviation of 0.012) more gross revenue on average compared to the control group. The lift of TOT is 1.428%, and the 95% CI is (0.633, 0.679), suggesting the lift of TOT on gross revenue is statistically significant.

**c. Measurement metric evaluation**

Pre-randomization check on ITT:

```
        Welch Two Sample t-test

data:  df$past_sales[df$treatment == 1] and df$past_sales[df$treatment == 0]
t = 0.7, df = 3e+06, p-value = 0.5
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0061  0.0133
sample estimates:
mean of x mean of y
    0.516     0.512
```

After conducting a pre-experiment randomization check on past sales, it turns out that only ITT yields a valid result (*p-value = 0.5, fail to reject the null*), meaning that treatment and control have similar past sales patterns. TOT yields a very small p-value that allows us to conclude that the randomization is not properly done (i.e., there is a significant difference in past sales before running the experiment), prohibiting us from using TOT as the ideal estimate. Since ITT is the only valid measure based on t.test, we should report ad's ITT of $894339 and a lift of 1.14% to the manager, recommending the effectiveness of the ad campaign due to statistical significance.

**d. The managerial and statistical implications of the metrics**

On average, the ITT of the ad campaign on total incremental gross revenue is $894339, meaning the treatment group in total generated much more gross sales compared to the control group. The lift of ITT on gross revenue is 1.14%, suggesting that the treatment group has 1.14% more gross revenue than the control group. Although we can't measure the ROI of the campaign, these statistics tell us that the ad is effective and generates statistically significant results.

## 6. Operation experiment

**a. Compute the observational estimate.**

```
        Welch Two Sample t-test

data:  df$sales[df$saw_ads == 1 & df$treatment == 1] and df$sales[df$saw_ads == 0 &
df$treatment == 1]
t = 58, df = 2e+06, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.642 0.687
sample estimates:
mean of x mean of y
    1.547     0.882

Observational estimated effect is: 0.665
Lift of effect is: 0.754
SD of effect is: 0.011
```

If we don’t have an experiment and control-group data, the point estimate for the observed effect of ad is $0.665, with a standard deviation of 0.011 and 95% CI of (0.642, 0.687). The estimated lift suggests that on average, the ad campaign generated 0.754% more sales for users who saw ads than those who didn’t.

**b. Possible issue**

If we only have observational estimates without control groups, this makes it impossible for us to study the effect on outcome variables by manipulating instrumental variables. The outcome could also be influenced by extraneous variables unrelated to experiments. Without control group and placebo ads, we can’t determine the accurate effect of the ad campaign on sales.

## 7. Market Segment Plan

**a. The average ad effect per UK customer?**

```
        Welch Two Sample t-test

data:  df$sales[df$treatment == 1 & df$IsUK == 1] and df$sales[df$treatment == 0 & df$IsUK
== 1]
t = 26, df = 952673, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.468 0.546
sample estimates:
mean of x mean of y
     1.72      1.21

ITT per UK customer is: 0.507
Lift of ITT per UK customer is: 0.419
SD of ITT is: 0.02
```

Point estimate of ad's ITT per non-UK customer is 0.507, with a standard deviation of 0.02 and lift of 0.419%. The p-value and 95% CI (0.468, 0.546) suggest the ad's TOT to be significant, so the ad campaign has a meaningful impact on a per-user basis in the UK.

**b. The average ad effect per customer outside of the UK?**

```
        Welch Two Sample t-test

data:  df$sales[df$treatment == 1 & df$IsUK != 1] and df$sales[df$treatment == 0 & df$IsUK
!= 1]
t = 36, df = 3e+06, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.310 0.345
sample estimates:
mean of x mean of y
    1.094     0.767

ITT per non-UK customer is: 0.327
Lift of ITT per non-UK customer is: 1.16
SD of non-UK TOT is: 0.009
```

Point estimate of ad's ITT per non-UK customer is 0.327, with a standard deviation of 0.009 and lift of 1.16%. The p-value and 95% CI (0.310, 0.345) suggest the ad's ITT to be significant, so the ad campaign has a meaningful impact on a per-user basis in non-UK. Although the ITT for non-UK users is smaller than UK users, the lift is higher (1.16% vs. 0.419%), meaning that users' sales response is more pronounced outside of the UK.

**c. Recommendation for allocating the budget**

I would recommend targeting more non-UK users during future ads campaigns due to its higher lift and more precise confidence intervals. Although non-UK user's ITT is $0.327 compared to $0.507 for UK users, there are much more non-UK observations (2,999,450 vs. 1,000,696) and a higher lift on ad campaigns can be converted to more sales as the ad reaches more users.