Assignment 2 Notebook

Shu Wang

RocketFuel

Remember to write out your answers in words, don't just output R statistics.

1.1 What is the ATE hat of the ads on conversion?

```
#calculate size of control& treatment group
ads_data%>% count(test)
```

```
   test      n
1:    0  23518
2:    1 564530
```

There are 588,048 observations. There are 564,530 observations in treatment group and 23,518 observations in control group.

```
aggregate(ads_data$converted, list(ads_data$test), FUN=sum)
```

```
  Group.1     x
1       0   420
2       1 14423
```

```
conversion_tr = 14423/564530
conversion_con = 420/23518
print(paste("converstion rate of treatment group is",conversion_tr))
```

```
[1] "converstion rate of treatment group is 0.0255486865179884"
```

```
print(paste("converstion rate of control group is",conversion_con))
```

```
[1] "converstion rate of control group is 0.0178586614508036"
```

```
conversion_tr = 14423/564530
conversion_con = 420/23518

print(paste("ATE(hat) of conversion is",conversion_tr-conversion_con))
```

```
[1] "ATE(hat) of conversion is 0.00769002506718477"
```

1.2 Did the campaign cause more purchases? Is this difference statistically significant?

μ_1 : conversion rate of treatment group(group with ads) μ_2 : conversion rate of control group(group with PSA)

Null hypothesis: $\mu_1 = \mu_2$ Alternative hypothesis: $\mu_1 \neq \mu_2$

```
#t.test(ads_data[test == 1, tot_impr], ads_data[test == 0, tot_impr])
```

Modify the function above to get the right answer. Your answer in the code chunk below

```
t.test(ads_data[test == 1, converted],ads_data[test == 0,converted])
```

```
    Welch Two Sample t-test

data:  ads_data[test == 1, converted] and ads_data[test == 0, converted]
t = 8.6523, df = 26376, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.005947966 0.009432084
sample estimates:
 mean of x  mean of y
0.02554869 0.01785866
```

We have a $p-$value < 0.05.At 5% significance level,we have strong evidence against the null hypothesis that $\mu_1 = \mu_2$

So we conclude that the difference is significant at 5% level.

2.2 Was the campaign profitable?

2.2.a How much more profit did TaskaBella make by running the campaign (excluding advertising costs)

Hint: the profit per conversion is given on page 2 of the case.

```
#Assuming the same conversion rate for treatment group and exposed group without campaign.
#The extra profit was gained from those users in exposed group that was converted due to ads
profit_camp = 564530 * 40 * (conversion_tr - conversion_con)
profit_camp
```

```
[1] 173650
```

2.2.b What was the cost of the campaign (including the control group)?

Hint: The cost per thousand impressions is $9

```
total_impr = sum(ads_data$tot_impr)
cost_camp = total_impr/1000 * 9
cost_camp
```

```
[1] 131385.8
```

2.2.c Calculate the ROI of the campaign (including the control group). Was the campaign profitable?

The ROI is calculated by (Effect on Profits per Person in Campaign - Cost of Ads per Person in Campaign) / (Cost of Ads per Person in Campaign)

```
ROI_camp = (profit_camp - cost_camp)/cost_camp
ROI_camp
```

```
[1] 0.3216801
```

2.2.d What was the opportunity cost of including a control group — how much more could TaskaBella have made by not having a control group at all?

```
oppr_cost = 23518 * (conversion_tr - conversion_con) * 40
oppr_cost
```

```
[1] 7234.16
```

3 Based on the above figure, can we say that more impressions cause more conversions? (No more than 2 sentences)

Based on the figure, more impressions shows more conversion rate within each group generally.

4 Calculate the power of this experiment.

4.1 Calculate cohen's D. Cohen's D in this case is the estimated average treatment effect on conversion divided by the standard deviation of conversion.

Hint, the standard deviation function is: sd

```
sd(ads_data$converted)
```

```
[1] 0.1568568
```

```
cohen_D = (conversion_tr - conversion_con)/sd(ads_data$converted)
cohen_D
```

```
[1] 0.04902576
```

4.2 Use the pwr.t2n.test function to calculate the power of the experiment:

Hint, we can calculate the number of individuals in a subset of the data like this: ads_data[test == 1, .N]

```
power_test <-pwr.t2n.test(n1 = ads_data[test == 1, .N],
n2 = ads_data[test == 0, .N],
d = cohen_D,
sig.level = .05,
power = NULL)
```

```
power_test
```

```
     t test power calculation

             n1 = 564530
             n2 = 23518
              d = 0.04902576
      sig.level = 0.05
          power = 1
    alternative = two.sided
```

```
#pwr.t2n.test(n1 = ads_data[test == 1, .N],
#n2 = 'put your number of control here',
# d = 'put cohens d here',
# sig.level = .05,
# power = NULL)
```

4.3 What would the power be instead if the true effect had a Cohen's D of .01?

Hint: Copy the above function and modify accordingly.

```
power_test <-pwr.t2n.test(n1 = ads_data[test == 1, .N],
n2 = ads_data[test == 0, .N],
d = 0.01,
sig.level = .05,
power = NULL)

power_test
```

```
     t test power calculation

             n1 = 564530
             n2 = 23518
              d = 0.01
      sig.level = 0.05
          power = 0.3239638
    alternative = two.sided
```

4.4 What would the power be instead if the true effect had a Cohen's of .01 and the sample was equally split between treatment and control?

Hint: Copy the above function and modify accordingly.

```
power_test <-pwr.t2n.test(n1 = 588048/2,
n2 = 588048/2,
d = 0.01,
sig.level = .05,
power = NULL)

power_test
```

```
     t test power calculation

             n1 = 294024
             n2 = 294024
              d = 0.01
      sig.level = 0.05
          power = 0.9695516
    alternative = two.sided
```

5 Case Writeup + Case Discussion in Class, be prepared to discuss!

Please write what you would discuss in your presentation to TaskaBella. Your answer should be **one** paragraph and the paragraph should be **five** or fewer sentences. Be prepared to discuss in class (part of the grade). Think about what is the most important thing to say to TaskaBella. No additional analysis is needed to answer this question.

Your answer here

From our above analysis,it is clear that the campaign on advertisement would lead to a higher conversion rate. At a reasonable control group size,the campaign is profitable and would lead to an increase in sales. There's a tradeoff between impression and cost.Namely,more impression would lead to higher conversion rate,but it also increases the cost.It is possible to find an optimized point which makes the highest profit.

How long did this assignment take you to do (hours)? How hard was it (easy, reasonable, hard, too hard)?

reasonable