Assignment 2 Notebook

Prateek Naharia

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?

Test Group = Receiving advertisements Control Group = Not receiving advertisements ATE hat, after splitting the data set in test & control, we subtract the mean of both the groups(test==0 & test==1) for 'converted'.

```
data_test <- ads_data[test == 1]
data_control <- ads_data[test == 0]
mean_converted_test <- mean(data_test$converted)
mean_converted_control <- mean(data_control$converted)
ATE_hat <- mean_converted_test - mean_converted_control
ATE_hat
```

```
[1] 0.007690025
```

Obtained ATE hat is 0.007690025.

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

Hint: Use the t.test function. For example, the code below conducts a t-test on the number of impressions. The p-value is .8274 and 95% CI is (-0.4972735 0.6217286).

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

```
    Welch Two Sample t-test

data:  ads_data[test == 1, tot_impr] and ads_data[test == 0, tot_impr]
t = 0.2168, df = 25602, p-value = 0.8284
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4977312  0.6215331
sample estimates:
mean of x mean of y
  24.8277   24.7658
```

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
```

Observation : 1. The p-value is less than 2.2e-16, p<0.05. 2. The 95% CI, also provides a range of values for the difference that is likely to contain the true value with 95% certainty. - Thus results that the ad campaign cause more purchases & is statistically significant.

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.

```
#$40 converting user worth * dim[converted ==1]
ads_data_test_con <- sum(ads_data$test == 1)
ads_data_control_con <- sum(ads_data$test == 0)

con_test <- sum(ads_data$converted[ads_data$test == 1])
con_control <- sum(ads_data$converted[ads_data$test == 0])

test_rate_conversion <- con_test/ads_data_test_con
control_rate_conversion <- con_control/ads_data_control_con
test_group = sum(ads_data$test == 1)
```

```r
more_profit <- (test_rate_conversion - control_rate_conversion) * test_group * 40
print(more_profit)
```

```
[1] 173650
```

```r
print(paste0("More Profit: ", more_profit))
```

```
[1] "More Profit: 173649.994047113"
```

More Profit appears to be exactly as 173649.99, rounding off as 173650. Ans 2.2 -> Yes, Profitable

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

Hint: The cost per thousand impressions is $9

```r
#14.5*10^6
cost_per_imp <- 9/1000
campaign_cost <- sum(ads_data[, tot_impr]) * cost_per_imp
print(paste0("The Cost of the campaign: ",campaign_cost))
```

```
[1] "The Cost of the campaign: 131385.798"
```

The Cost of the campaign appears to be $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)

```r
roi <- (more_profit - campaign_cost)/campaign_cost
print(paste0("The ROI of the campaign is: ", roi))
```

```
[1] "The ROI of the campaign is: 0.321680095493372"
```

The ROI appears to be +ve as 0.32168 or 32%, hence is profitable.

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?

```
data_test = sum(ads_data$test == 1) #//reference from question 1.1
data_control = sum(ads_data$test == 0) #// reference from question 1.1

conv_rate_treatment <- sum(ads_data$converted[ads_data$test == 1]) / data_test
conv_rate_control <- sum(ads_data$converted[ads_data$test == 0]) / data_control

control_group <- sum(ads_data$test == 0)

opportunity_cost <- (conv_rate_treatment - conv_rate_control) * control_group * 40
print(paste0("The Opportunity Cost of including a control group is:  ",opportunity_cost))
```

```
[1] "The Opportunity Cost of including a control group is:  7234.16038120206"
```

Opportunity Cost : 7234.16038120206

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

Ans Q3: No,Based on the above figure we cannot say as its directly proportional, but correlated also, The causality cannot be established without considering the assumptions of casual inference.

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

```
test_data <- ads_data[test == 1, converted]
control_data <- ads_data[test == 0, converted]

mean_data_test <- mean(test_data)
mean_data_control <- mean(control_data)

standard_deviation_control = sd(control_data)
cohen_d = (mean_data_test - mean_data_control)/standard_deviation_control
cohen_d
```

```
[1] 0.05806402
```

```
#cohens_d = ATE_hat/standard_deviation_control
#cohens_d / Alternatively for results (cohen's d)

print(paste0("The value of mean of test group: ", mean_data_test))
```

```
[1] "The value of mean of test group: 0.0255486865179884"
```

```
print(paste0("The value of mean of control group: ", mean_data_control))
```

```
[1] "The value of mean of control group: 0.0178586614508036"
```

```
print(paste0("The value of standard deviation of control group: ", standard_deviation_contro
```

```
[1] "The value of standard deviation of control group: 0.132440460184666"
```

```
print(paste0("The value of cohen's d: ", cohen_d))
```

```
[1] "The value of cohen's d: 0.0580640165132491"
```

Cohen's D value is 0.0580640165132491 Standard Deviation appears to be: 0.132440460184666

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]

```
pwr.t2n.test(n1 = sum(ads_data[test == 1, .N]),
n2 = sum(ads_data[test == 0, .N]),
d = cohen_d,
sig.level = .05,
power = NULL)
```

```
     t test power calculation

             n1 = 564530
             n2 = 23518
              d = 0.05806402
      sig.level = 0.05
          power = 1
    alternative = two.sided
```

Here in above treatment with moderate effect size, the power appears to be 1 or 100% probability of detecting an effect if exists.

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.

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

```
     t test power calculation

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

power = 0.3239638: This states, experiment has a 32.4% chance of detecting the true effect if it exists, Also it will consider both positive and negative differences between treatment and control group.

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.

```
size = ads_data[,.N]/2
pwr.t2n.test(n1 = size,
             n2 = size,
             d = 0.01,
             sig.level = .05,
             power = NULL)
```

```
     t test power calculation

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

Power: 0.9695516, or approx 97%.

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

I would discuss the real-world examples & importance of online advertising and its impact on the business. Having opportunity ($7234) for including control group & +ve ROI, does holds a scope of online advertising. It is also seen that, an increase profit of $173650 is enough for advertising campaign to be considered.

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

The assignment took a day, & hard.