Assignment 2

Chiebuka Onwuzurike

RocketFuel

Read the Data:

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.5     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

2.1 Did the campaign cause more purchases?

Is this difference statistically significant?

Yes, the campaign caused more purchases. There is a difference of 0.00769245 between the mean of those who saw the campaign and those who didn't. With as CI of (0.005950817 - 0.009434089 and a p-value of 2.2e-16, we are 95% confident that there is an effect.

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

```
##
##  Welch Two Sample t-test
##
## data:  ads_data[test == 1, converted] and ads_data[test == 0, converted]
## t = 8.6572, df = 26384, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.005950817 0.009434089
## sample estimates:
##  mean of x  mean of y
## 0.02554656 0.01785411
```

2.2 Was the campaign profitable?

2.2.a How much more profit did TaskaBella make by running the campaign (excluding advertising costs) (8 points)

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

Yes the campaign was profitable. With an average treatment effect (ATE) of 0.007692453 the profit if everyone was shown the add would be $180,957.6 dollars.

```
profit_ad = ttest$estimate[1] * ads_data[,.N] * 40
profit_noad = ttest$estimate[2] * ads_data[,.N] * 40
# OR
ATE = ttest$estimate[1] - ttest$estimate[2]
profit_ate = ATE * ads_data[,.N] * 40

profit_per_person = profit_ate/ads_data[,.N]
profit_ate
```

```
## mean of x
##  180957.6
```

2.2.b What was the cost of the campaign? (7 points)

Hint: The cost per thousand impressions is $9

The cost of the campaign was $131,374.6.

```
ad_spend_total = sum(ads_data$tot_impr)/1000*9
ad_spend_per_person = ad_spend_total/ads_data[,.N]
ad_spend_total
```

```
## [1] 131374.6
```

2.2.c Calculate the ROI of the campaign. Was the campaign profitable? (7 points)

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

With a profit per person of 0.3076981 and a cost per person of 0.2233879, the ROI is 0.3774164.

```
ROI = (profit_per_person -ad_spend_per_person)/ad_spend_per_person
ROI
```

```
## mean of x
## 0.3774164
```

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? (7 points)

If the 23,524 users were not test with the control advertisement they would have been more likely to convert. This 4% control group was an opportunity cost of 7,238.29 for TaskaBella.

```
opp_cost = (ATE * ads_data[test == 0,.N] * 40)

opp_cost
```

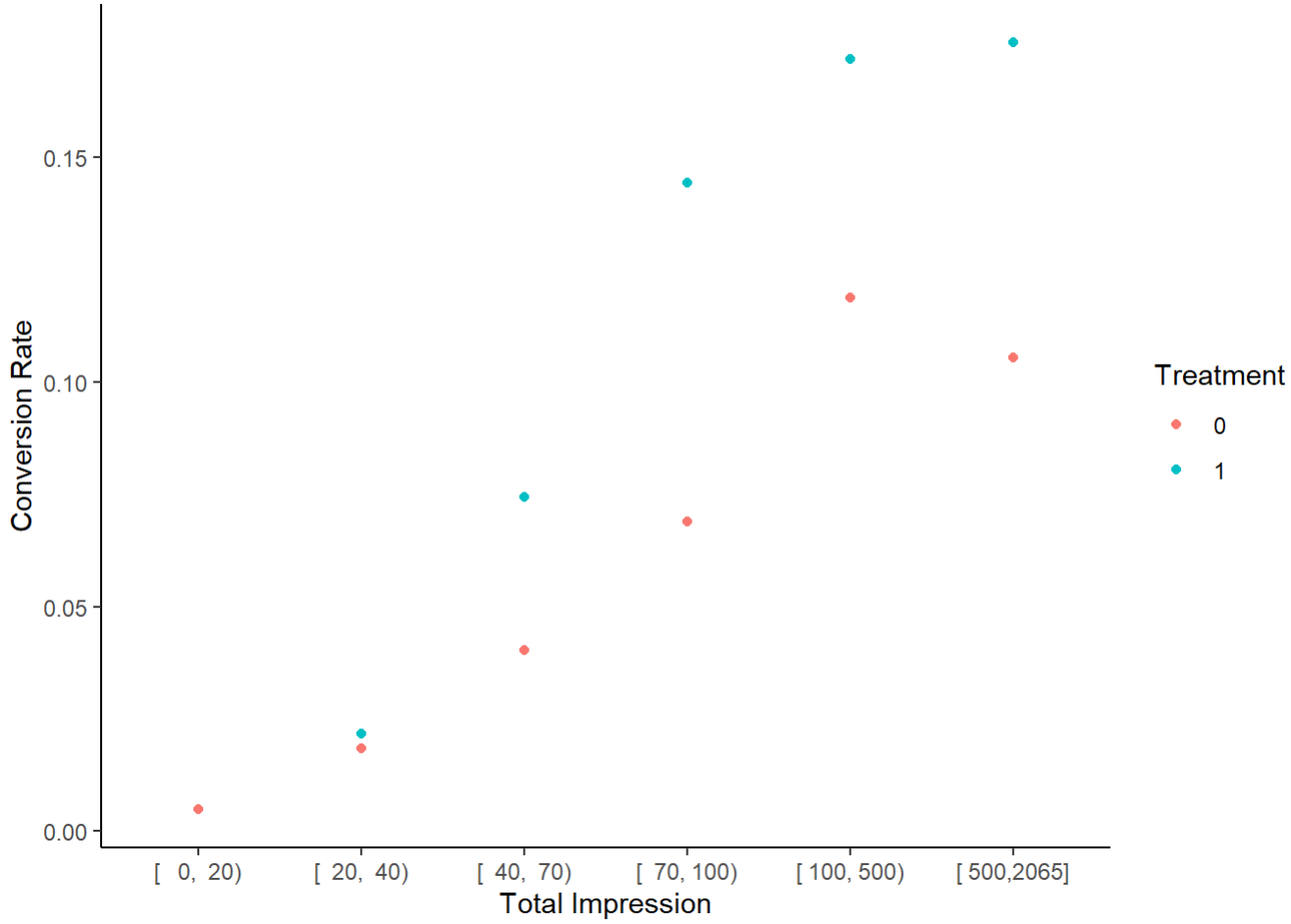
```
## mean of x
##  7238.291
```

3 Did the number of impressions seen by each user influence the effectiveness of advertising?

3.a Create a chart of conversion rates as a function of the number of ads displayed to users. Plot conversion rates for those who were in the control group and for those who were exposed to the (7 points)

```
### Create bins of the variable 'tot_impr'
ads_data[, group_tot_impr := cut2(tot_impr, c(0,20,40,70,100,500))]

aggregate_by_num <- ads_data[, list(conversion_rate = mean(converted), num_user = .N), by = list
(group_tot_impr, test)]
this_plot <- ggplot(aggregate_by_num, aes(x = group_tot_impr, y = conversion_rate, color = facto
r(test))) + geom_point() + xlab("Total Impression") + ylab("Conversion Rate") + theme_classic()
 + labs(color="Treatment")
this_plot
```




```
aggregate_by_num
```

```
##     group_tot_impr test conversion_rate num_user
##  1:    [ 100, 500)    1     0.171795693    21869
##  2:    [  70, 100)    1     0.144159334    18979
##  3:    [  20,  40)    1     0.021735350   109269
##  4:    [ 500,2065]    1     0.175438596      570
##  5:    [   0,  20)    1     0.004820131   364098
##  6:    [  40,  70)    1     0.074309126    49792
##  7:    [ 100, 500)    0     0.118694362     1011
##  8:    [  20,  40)    0     0.018547141     4529
##  9:    [   0,  20)    0     0.004965572    15104
## 10:    [  70, 100)    0     0.068840580      828
## 11:    [  40,  70)    0     0.040334481     2033
## 12:    [ 500,2065]    0     0.105263158       19
```

3.b What can you infer from the charts? For what conversion number was advertising most effective? (7 points)

One can infer a positive linear relationship between the number of advertisement displayed to a user and the conversion rate. Impressions between 500 and 2065 resulted in the highest conversion rate. It should be note that only 570 treatment group users and only 19 control group users received impressions in this range.

3.c Based on the above figure, suggest a follow-up experiment. What is the treatment and what is the control? Answer should be 1 paragraph. (8 points)

The first experiment concluded in that the campaign is in fact effective. Let's say a large part of the CPM is the size of the add. A follow-up experiment would be to see what size of the ad is the optimal size given the increase/decrease in cost and effective. The control would be the add at the current size and the treatment would be a smaller ad.

4 Calculate the power of this experiment.

4.1 Calculate cohen's D. This is the treatment effect on conversion divided by the standard deviation of conversion.

Hint, the standard deviation function is: sd With an standard deviation for conversion of 0.1568 and and ATE of 0.007692, Cohen's D is 0.04904.

```
sd_conversion = sd(ads_data$converted)
sd_conversion
```

```
## [1] 0.1568499
```

```
cohensd = ATE/sd_conversion
cohensd
```

```
##  mean of x
## 0.04904339
```

4.2 Use the pwr.t2n.test function to calcualte 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]

The power is 1.

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

```
##
##      t test power calculation
##
##              n1 = 564577
##              n2 = 23524
##               d = 0.04904339
##       sig.level = 0.05
##           power = 1
##     alternative = two.sided
```

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.

The power would be 0.3240307.

```
pwr.t2n.test(n1 = ads_data[test == 1, .N], n2 = ads_data[test == 0, .N], d = .01, sig.level = .0
5, power = NULL)
```

```
##
##      t test power calculation
##
##              n1 = 564577
##              n2 = 23524
##               d = 0.01
##       sig.level = 0.05
##           power = 0.3240307
##     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.

The power would be 0.9695635.

```
pwr.t2n.test(n1 = ads_data[, .N]/2, n2 = ads_data[, .N]/2, d = .01, sig.level = .05, power = NUL
L)
```

```
##
##      t test power calculation
##
##              n1 = 294050.5
##              n2 = 294050.5
##               d = 0.01
##       sig.level = 0.05
##           power = 0.9695635
##     alternative = two.sided
```

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

I would first start off the meeting with TaskaBella and congratulate them on a profitable and successful product because it brough in 593,720 dollars worth of profit. Then I would have shifted the discussion to how much was due to the ads and how much was due to the great product and word of mouth. I would make a say that around 30% of the profit was due to our ads which equals 180,957.6 dollars. I would explain how the average conversion rate of the control group that without seeing our ads was about 1.78%. However the conversion rate for the treatment group that saw the ads was about 2.55%; a 43% increase. I would then drive home the point by talking about the 37.74% ROI given the fact the profit per person was 0.3076981 and a cost per person of 0.2233879. They would obviously be a little skeptical, and want us to back up our points as we promised. I would restate how we didn't show the ads to 4% of our target group, which did have an opportunity cost 7,238.29 dollars of profit. However I would drive home the fact that this cost was necessary to be 95% confident that our ads did have effect we are claiming. For the engineers in the room I would highlight the p-value of 2.2e-16 and our statistical power of 1 given our Cohen's D of 0.04904339. Being the good salesman I am, I would finally shift the conversation to follow up ads and experiments. I would talk about how we can do a follow experiment to see what size ad is most effective and efficeint.

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

On and off for a few days but I would say a total of 4-6 hours. For me it took a very long time wrapping my mind around what was being asked of me. I don't know, it felt like the jump from our theoretical "conversations" about the topics during lectures, to actual applications of the material was quite large. Question 2 and slightly 3 took the bulk of the time.