Assignment 2 Notebook

Yuxuan Mei

RocketFuel

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

1.1 What is the ATE of the ads on purchases? (4 points)

```
treat_cr <- ads_data[(test==1) & (converted==1),.N ] / ads_data[(test==1),.N ]
control_cr <-ads_data[(test==0) & (converted==1),.N ] / ads_data[(test==0),.N ]

ATE <- treat_cr - control_cr
ATE
```

```
## [1] 0.007692453
```

1.2 Did the campaign cause more purchases? Is this difference statistically significant? (6 points)

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])
```

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 = 8.6572, 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
```

The t-test statistics shows that the ads did have a statiscticall significant impact on consumer purchases as the p-value is very low. ### 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.

```
profit <- ads_data[(test == 1), .N ] * ATE *40
profit
```

```
## [1] 173719.3
```

We would expect the campaign to be profitable.

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

Hint: The cost per thousand impressions is $9

```
cost <- sum(ads_data[,tot_impr])/1000 * 9
cost
```

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

The cost of the campaign is expected to be $131375.

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

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 <- (profit-cost)/cost
ROI
```

```
## [1] 0.3223198
```

The campaign is expected to be profitable with an ROI of 32%.

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)

```
opportunity_cost <- ATE * ads_data[(test == 0),.N ]*40
opportunity_cost
```

```
## [1] 7238.291
```

The opportunity cost is expected to be $7238.

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

Group together number of impressions as necessary to obtain a meaningful plot. (Conversion rate means the percentage of unique users who made a purchase.) (7 points)

Below is an example of a similar plot where the outcome is impressions per day. Change the values to answer 3.a and label the plot.

```
library(Hmisc)
```

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

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

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

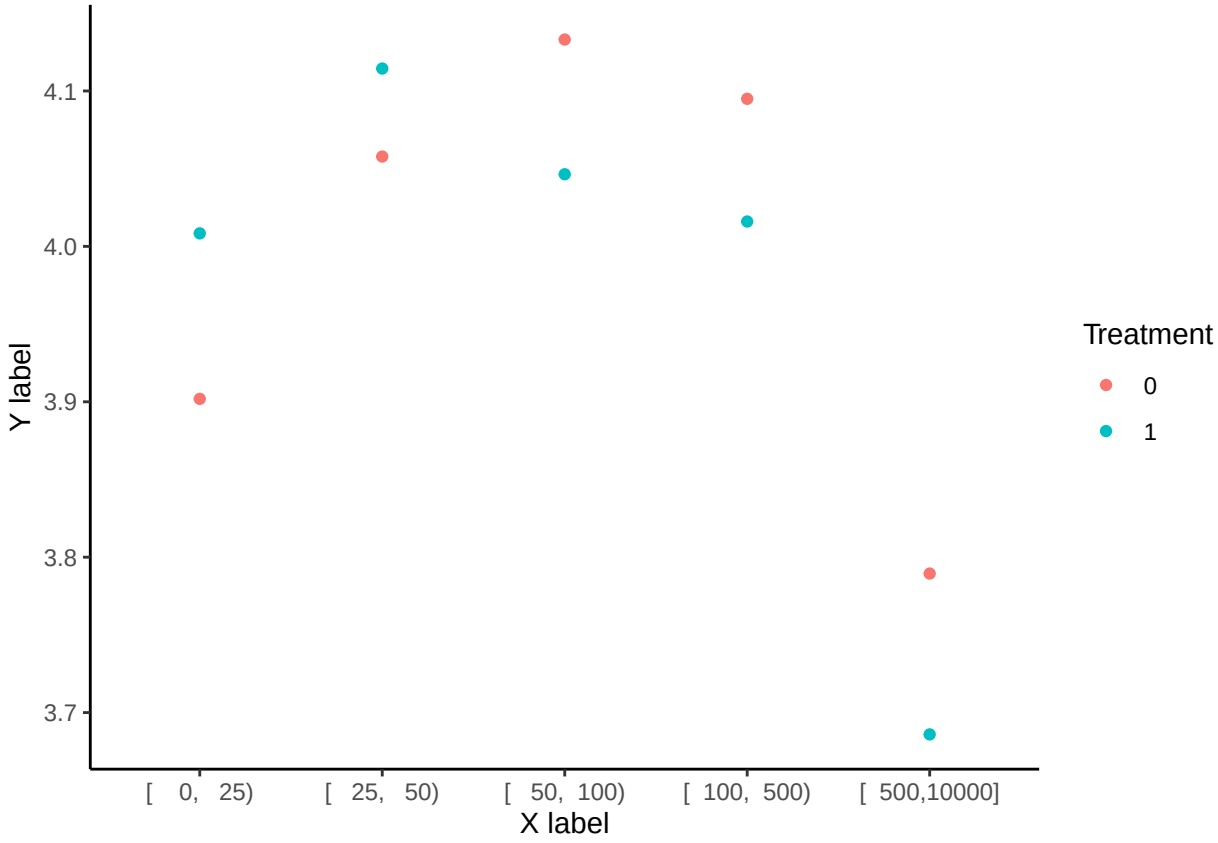
```
##
## Attaching package: 'Hmisc'
```

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

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

```
### Create bins of the variable 'tot_impr'
ads_data[, group_tot_impr := cut2(tot_impr, c(0, 25, 50, 100, 500, 10000))]

aggregate_by_num <- ads_data[, list(mean_impr_day = mean(mode_impr_day)), by = list(group_tot_impr, tes
this_plot <- ggplot(aggregate_by_num, aes(x = group_tot_impr, y = mean_impr_day, color = factor(test)))
this_plot
```

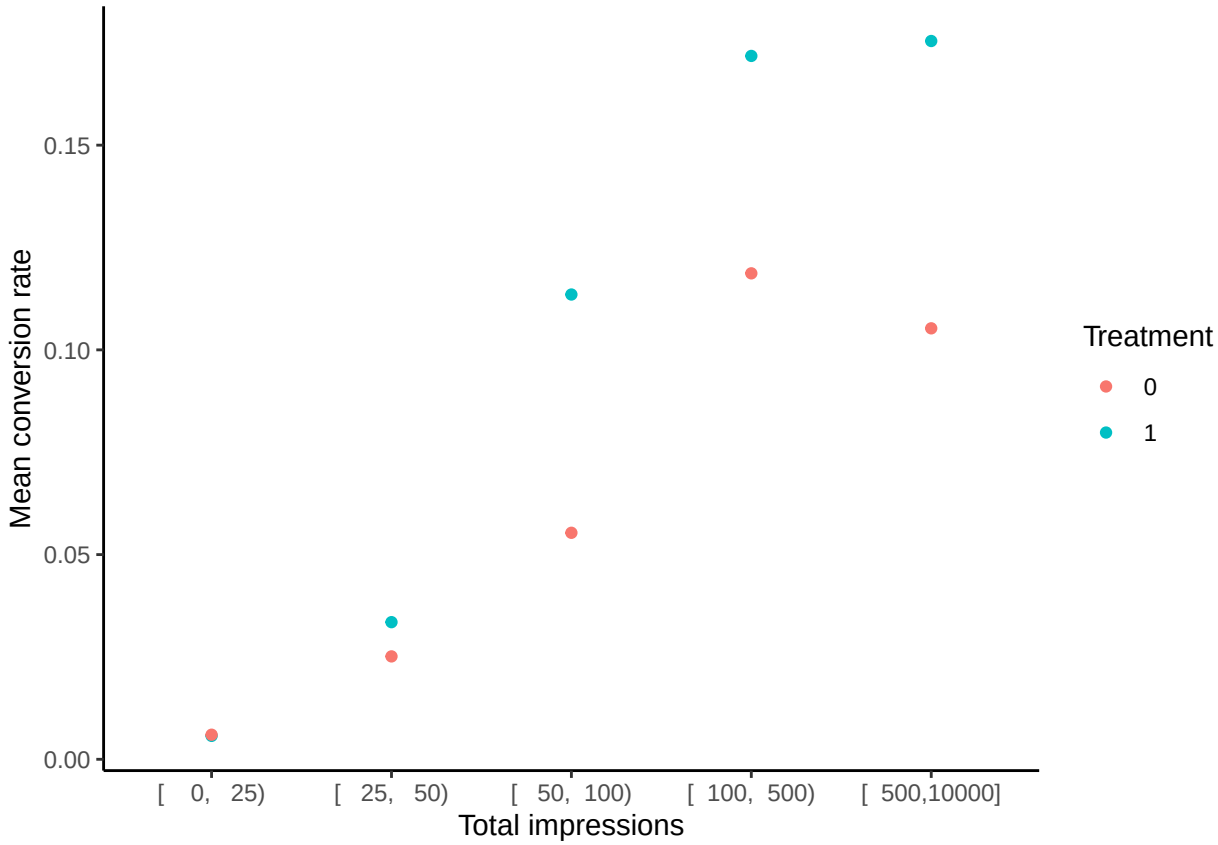



```
ads_data[, group_tot_impr := cut2(tot_impr, c(0, 25, 50, 100, 500, 10000))]
aggregate_by_num <- ads_data[, list(conv_rate = mean(converted)), by = list(group_tot_impr, test)]

#aggregate_by_num <- ads_data[(converted == 1), .N, .(group_tot_impr, test)]



this_plot <- ggplot(aggregate_by_num, aes(x = group_tot_impr, y = conv_rate, color = factor(test))) +
  geom_point() + xlab("Total impressions") + ylab("Mean conversion rate") +
  theme_classic() + labs(color="Treatment")
this_plot
```



3.b Based on the above figure, can we say that more impressions cause more conversions? (No more than 2 sentences) (8 points)

Write answer here.

The graph suggests that more impressions is positively correlated with more conversions. However, we can't say there's a causality effect because the number of total impressions is not randomized and we can only conclude correlation.'

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

```
cohen_d <- ATE/ads_data[, sd(converted)]
cohen_d
```

```
## [1] 0.04904339
```

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 = 'put your number of treated here', n2 = 'put your number of control here', d = 'put
```

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

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

Power of the experiment is 1.

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 = ads_data[test==1, .N] , n2 = ads_data[test==0, .N], d = 0.01, sig.level=.05, power=NU
```

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

Power would be 0.3240307.

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.

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

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

Power would be 0.9695635 if the sample was split equally. ### 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 2 paragraphs. Be prepared to discuss in class (part of the grade). Each paragraph should be 5 or fewer sentences. Think about what is the most important thing to say to TaskaBella.

Our experiment shows that running advertisements is expected to have a statistically significant ROI of 32% (really high). Generally, more advertisement is positively correlated with higher conversion rates. However, we need to be careful when setting the number of impressions to show, as there could be have decreasing margin of impact when the number of impressions increases. So overall speaking, we would recommend implementing the advertisements as it is expected to help us generate profit.

Also, there's a trade off between control group size, statistical power and opportunity cost. A larger control group size would contribute in having a higher statistical power, but it will have more opportunity cost for the company. We also recommend trying different treatments by having different advertisements to see which type of advertisements could be more effective.

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

3 hours; it's reasonable