causal_experiment

Chieh Hsin(Jessica) Cheng

2023-09-26

(a) If the control and treatment groups are similar across tenure, premium_user, and num_posts_before metrics.

(a-1) tenure

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
data_1 <- read.csv('data_Q1.csv')
t.test(tenure ~ treated, data_1)
```

```
##
##  Welch Two Sample t-test
##
## data:  tenure by treated
## t = 1.373, df = 1789.6, p-value = 0.1699
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -19.09774 108.23144
## sample estimates:
## mean in group 0 mean in group 1
##        572.1680        527.6011
```

Based on p-value: 0.1699, which is greater than the typical significance level 0.05, we do not reject null hypothesis. Hence, there are no significant difference between two groups.

(a-2) premium_user

```
t.test(premium_user ~ treated, data_1)
```

```
##
##  Welch Two Sample t-test
##
## data:  premium_user by treated
## t = 0.95906, df = 1769.9, p-value = 0.3377
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.006928414  0.020188082
## sample estimates:
## mean in group 0 mean in group 1
##      0.02541436      0.01878453
```

Based on p-value: 0.3377, which is greater than the typical significance level 0.05, we do not reject null hypothesis. Hence, there are no significant difference between two groups.

(a-2) num_post_before

```
t.test(num_post_before ~ treated, data_1)
```

```
##
##  Welch Two Sample t-test
##
## data:  num_post_before by treated
## t = 0.56253, df = 1796.1, p-value = 0.5738
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.2307971  0.4164325
## sample estimates:
## mean in group 0 mean in group 1
##        1.643094        1.550276
```

Based on p-value: 0.5738, which is greater than the typical significance level 0.05, we do not reject null hypothesis. Hence, there are no significant difference between two groups.

(b) Does getting reddit gold increase likelihood that the user will post (use the posted metric as the dependent variable and treated as the independent variable)? Use a simple linear model (not a logit) for the analysis

```
summary(lm(posted ~ treated, data_1))
```

```
##
## Call:
```

```
## lm(formula = posted ~ treated, data = data_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6232 -0.5602  0.3768  0.4398  0.4398
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.56022    0.01631   34.34   <2e-16 ***
## treated      0.06298    0.02307    2.73   0.0064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4908 on 1808 degrees of freedom
## Multiple R-squared:  0.004105,   Adjusted R-squared:  0.003554
## F-statistic: 7.452 on 1 and 1808 DF,  p-value: 0.006396
```

Posted = 0.55 + 0.063*treated

0.55 is the mean of group 0 (control group) and 0.063 is the difference two groups

Look at p-value: 0.0064, which is smaller than 0.05, meaning that the feature made significant difference

(c) What sorts of users are more likely to increase their contribution? (use the tenure and the first_timer variables)

```
summary(lm(posted ~ first_timer * treated, data_1))
```

```
##
## Call:
## lm(formula = posted ~ first_timer * treated, data = data_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6370 -0.6120  0.3630  0.3880  0.5031
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.630841   0.023621  26.706  < 2e-16 ***
## first_timer         -0.133986   0.032536  -4.118 3.99e-05 ***
## treated              0.006196   0.033877   0.183   0.8549
## first_timer:treated  0.108949   0.046107   2.363   0.0182 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4887 on 1806 degrees of freedom
## Multiple R-squared:  0.01369,    Adjusted R-squared:  0.01205
## F-statistic: 8.354 on 3 and 1806 DF,  p-value: 1.623e-05
```

Here I mainly focus on the interaction term(first_timer and treated). Look at p-value: 0.0182, which is smaller than 0.05. This suggests that the relationship between "first_timer*treated" and the likelihood of posting is statistically significant based on the results.

```
summary(lm(posted ~ tenure * treated, data_1))
```

The coefficient of interactive term is 0.109 suggests that the interaction has a positive effect on contributions. Meaning that first-time users who are treated tend to contribute more.

```
##
## Call:
## lm(formula = posted ~ tenure * treated, data = data_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6548 -0.5599  0.3477  0.4382  0.5396
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.585e-01  2.077e-02  26.893  < 2e-16 ***
## tenure          2.934e-06  2.250e-05   0.130  0.89627
## treated         9.622e-02  2.949e-02   3.263  0.00112 **
## tenure:treated -6.275e-05  3.357e-05  -1.869  0.06174 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4903 on 1806 degrees of freedom
## Multiple R-squared:  0.007284,   Adjusted R-squared:  0.005635
## F-statistic: 4.417 on 3 and 1806 DF,  p-value: 0.004211
```

Here I mainly focus on the interaction term(tenure and treated). Look at p-value: 0.06174, which is greater than 0.05. This suggests that the relationship between "tenure*treated" and the likelihood of posting is not statistically significant based on the results. Meaning that the inteactive term has no strong relationship with the posted.

Also, the coefficient of the interactive term -6.275e-05 suggests that the interaction has a negative effect on contributions. Meaning that users with longer tenure who are treated tend to contribute slightly less.

In conclusion, users who are first timer user and treated with Reddit gold would be mroe likely to increase their contribution.

(d) Is the SUTVA assumption likely to be violated in the experiment?

- Interactions among users: Reddit has a compex ecosystem with various subreddits, user communities, and dynamics. Changes in behavior in one part of the platform can potentially spikk over or affect the behavior of uses in other parts of the platform.

Q2

(a) Use a t-test to see if there is a statistical difference in the pre-period between schools in the treatment (bal = 1) and control (bal = 0). This will check if randomization has been done correctly. To do this, calculate the average normalized test score(norm) for the pre period (pre = 1) for math (test_type = 0). Is there a statistical difference between students who got the Balsakhi program and did not get the program? Perform the same test for language (test_type = 1).

```
data_2 <- read.csv('data_Q2.csv')
pre_math_data <- subset(data_2, pre == 1 & test_type == 0 & std == 3)

t.test(norm ~ bal, data= pre_math_data)
```

```
##
##  Welch Two Sample t-test
##
## data:  norm by bal
## t = -1.0045, df = 5124.3, p-value = 0.3152
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.08207922  0.02646252
## sample estimates:
## mean in group 0 mean in group 1
##   -1.744781e-08    2.780833e-02
```

```
pre_language_data <- subset(data_2, pre == 1 & test_type == 1 & std == 3)

t.test(norm ~ bal, data= pre_language_data)
```

Based on p-value: 0.3152, which is larger than the typical significance level 0.05, we don't reject null hypothesis. Hence, there are no significant difference between two groups.

```
##
##  Welch Two Sample t-test
##
## data:  norm by bal
## t = -3.029, df = 5120.7, p-value = 0.002466
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.14426621 -0.03089769
## sample estimates:
## mean in group 0 mean in group 1
##    5.710910e-09    8.758195e-02
```

Based on p-value: 0.002466, which is smaller than the typical significance level 0.05, we do reject null hypothesis. Hence, there are significant difference between two groups.

```r
post_math_data <- subset(data_2, post == 1 & test_type == 0 & std == 3)

t.test(norm ~ bal, data= post_math_data)
```

(b) Calculate the average test scores for the post period (post = 1) for math for treatment and control. Is there a statistical difference between students in the two groups of schools? Use a ttest model to test the increase. Perform the same analysis for language test scores.

```
##
##  Welch Two Sample t-test
##
## data:  norm by bal
## t = -5.2818, df = 4221.1, p-value = 1.343e-07
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.2475276 -0.1135148
## sample estimates:
## mean in group 0 mean in group 1
##       0.2535332       0.4340544
```

```r
post_language_data <- subset(data_2, post == 1 & test_type == 1 & std == 3)

t.test(norm ~ bal, data= post_language_data)
```

Based on p-value: 1.343e-07, which is smaller than the typical significance level 0.05, we do reject null hypothesis. Hence, there are significant difference between two groups after the program.

```
##
##  Welch Two Sample t-test
##
## data:  norm by bal
## t = -4.2688, df = 4227.1, p-value = 2.008e-05
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.23207301 -0.08599442
## sample estimates:
## mean in group 0 mean in group 1
##       0.7151769       0.8742106
```

Based on p-value: 2.008e-05, which is smaller than the typical significance level 0.05, we do reject null hypothesis. Hence, there are significant difference between two groups.

c) Can you conclude if the Balsakhi program increase test scores in reading and mathematics?

Before the program implementation, there are no difference between two groups in mathematics. However, after the program, there are significant difference between two groups in math test. Hence, we could conclude that the program indee increase test scors in mathematics when others conditions were fixed.

When it comes to reading test(language), there are difference between two groups before the program. After the program, the p-value is extermely low, much less than 0.05, indicating a highly significant difference in test scorees between the two groups. Therefore, it appears that the Balskhi program has increased test scores. However, we should be cautious when making decision based on this result.

d) Is the SUTVA assumption viloated in the example?

Yes. The assumption might be viloated in the example.

- Inference between students: if students in the treatment group shared resources or knowledge with students in the control group, it could affect the control group's outcomes.

- Spillover Effect: If the presence of the Balsakhi program in one school had spillover effects on nearby schools or classrooms, leading to unintened consequences on test scores, it could violate the SUTVA assumption.