Casual Analysis

Claire Tsao

2023-09-23

```r
library(readxl)
library(dplyr)
```

```
##
## 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
```

```r
Reddit <- read.csv("data_Q1.csv")
```

```r
library(readxl)
library(dplyr)
Balsakhis <- read.csv("data_Q2.csv")
```

Q1

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

- tenure: p-value = 0.1699(>0.05), fail to reject H0. it has significant evidence that there is no different between treatment and control group. premium_user: p-value = 0.3377(>0.05), fail to reject H0, it has significant evidence that there is no different between treatment and control group.
- num_posts_before: p-value = 0.5738(>0.05), fail to reject H0, it has significant evidence that there is no different between treatment and control group.

```r
t.test(tenure ~ treated, Reddit)
```

```
##
##  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
```

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

```
##
##  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
```

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

```
##
##  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
```

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. P-value $= 0.0064 (<0.05)$, reject H0. it has significant evidence that there's a different between the treatment and control group. people who are getting reddit gold(treatment) have a 0.06298 higher chance of posting compared to those in the control group (didn't get Reddit Gold).

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

```
##
## Call:
## lm(formula = posted ~ treated, data = Reddit)
##
## 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
```

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

- The P-value of interaction effect between the first time users and whether being treated is 0.0182($<$0.05), reject H0.   This means that there is statistical significant that the interaction between being "treated" and being a "first_timer" has a effect on user contributions. For the first time users who are getting reddit gold(treatment), it has a statistically significant increase in user contributions(posted) which approximately equal to 0.108949 units.

- The P-value of interaction effect between the tenure and whether being treated is 0.06174($>$0.05), fail to reject H0. While there is a slight negative association between this interaction and contributions, it is not strong enough to be considered statistically significant in this analysis.

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

```
##
## Call:
## lm(formula = posted ~ treated * first_timer, data = Reddit)
##
## 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 ***
## treated             0.006196   0.033877   0.183   0.8549
## first_timer        -0.133986   0.032536  -4.118 3.99e-05 ***
## treated:first_timer 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
```

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

```
##
## Call:
## lm(formula = posted ~ treated * tenure, data = Reddit)
##
```

```
## 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 ***
## treated         9.622e-02  2.949e-02   3.263  0.00112 **
## tenure          2.934e-06  2.250e-05   0.130  0.89627
## treated:tenure -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
```

d) Is the SUTVA assumption likely to be violated in the experiment? It is possible that it exsits interaction between users who received Reddit Gold and those who did not.Users who receiving Reddit Gold might want to giving Reddit Gold to others.

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).

- Math: p-value = 0.3152(>0.05), fail to reject H0. This suggests that there is no statistically significant difference in the pre-period average normalized math scores (norm) between schools in the treatment group (bal = 1) and the control group (bal = 0).

- Language: p-value = 0.002466(<0.05), reject H0. This suggests that there is statistically significant difference in the pre-period average normalized language scores (norm) between schools in the treatment group (bal = 1) and the control group (bal = 0).

```
pre_math_df <- subset(Balsakhis, std ==3 & pre == 1 & test_type == 0)
t.test(norm ~ bal,pre_math_df)
```

```
##
## 	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_df <- subset(Balsakhis, std ==3 & pre == 1 & test_type == 1)
t.test(norm ~ bal,pre_language_df)
```

```
##
##  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
```

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 t- test model to test the increase. Perform the same analysis for language test scores.

- Math: p-value = 1.343e-07($<$0.05), reject H0. This suggests that there is a statistically significant difference in the post-period average normalized math scores (norm) between schools in the treatment group (bal = 1) and the control group (bal = 0).

- Language: p-value = 2.008e-05($<$0.05), reject H0. This suggests that there a is statistically significant difference in the post-period average normalized language scores (norm) between schools in the treatment group (bal = 1) and the control group (bal = 0).

```
post_math_df <- subset(Balsakhis, std ==3 & post == 1 & test_type == 0)
t.test(norm ~ bal,post_math_df)
```

```
##
##  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
```

```
post_language_df <- subset(Balsakhis, std ==3 & post == 1 & test_type == 1)
t.test(norm ~ bal,post_language_df)
```

```
##
##  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
```

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

- For Math, after the Balsakhi program, there is strong statistical evidence to suggest that the program has a positive and statistically significant effect on math test scores, with the group receiving the program (group 1) showing higher scores compared to the group without the program (group 0).

- For Language, after the Balsakhi program, there is a statistically significant difference in reading test scores, with the group receiving the program (group 1) maintaining higher scores compared to the group without the program (group 0). However, since the t-test before the Balsakhi program suggests a score difference exists between the treatment and control group, we need to be cautious when making the conclusion.

d)Is the SUTVA assumption violated in the example? Interference could occur between treatment and control group. For example, if the Balsakhi program has a positive effect on the weakest students and improves their academic performance, those students may become more engaged and motivated. As a result, when they return to the regular classroom, their improved performance might positively influence or "spill over" to the other students. The non-tutored students might benefit indirectly from the program even though they did not receive tutoring themselves.