Causal Inference: Assignment 1

Abhishikth Peri (5907987)

2023-09-30

Q1) Platforms use various methods to stimulate user's content creation. This includes paying users for reviews and providing awards and badges to users. Reddit is one of the largest platforms for creating and sharing content. On Reddit, users can recognize other contributions by providing gold to each other. However, does getting Reddit gold actually increase the receiver's content generation? To find out, researchers gave 905 random users reddit gold. Data is included for a similar number of users in the control group who did not receive gold during the time of the experiment. Import the data and examine:

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

```
library(readr)
```

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

```
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
```

```
Q1 = read_csv("data_Q1.csv")
```

```
## Rows: 1810 Columns: 22
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (22): user_id, treated, intervention_post_score, tenure, premium_user, c...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
t.test(tenure ~ treated, data = Q1)
```

```
##
##  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, data = Q1)
```

```
##
##  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, data = Q1)
```

```
##
##  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
```

In all of the three above cases, the p-value for t-test between columns tenure, premium_user and num_post_before with column treated are 0.1699, 0.3377 and 0.5738 respectively. Since p-value is greater

than 0.05 (level of significance) for all the cases, we can say that there is no statistical difference between tenure, premium_user and num_post_before metrics of treatment group and control group. Hence, there is similarity between treatment and control groups across tenure, premium_user and num_post_before metrics.

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.

```
M1 = lm(posted ~ treated, data = Q1)
summary(M1)
```

```
##
## Call:
## lm(formula = posted ~ treated, data = Q1)
##
## 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
```

Since the p-value from the hypothesis test (0.006396) < 0.05 (level of significance), there is evidence to suggest that providing Reddit Gold has a statistically significant effect on increasing the likelihood that a user will post.

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

```
M2 = lm(posted ~ treated * tenure , data = Q1)
summary(M2)
```

```
##
## Call:
## lm(formula = posted ~ treated * tenure, data = Q1)
##
## 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
```

Here, we observe that the interaction coefficient between treated and tenure is negative (-6.275e-05). However, we also observe that the p-value for this interaction (0.06174) is less than 0.05 (level of significance), which means that it is statistically similar to 0 and there is no strong statistical evidence to suggest that there is interaction between tenure and treated. Hence, tenure metric (number of days elapsed after user registered on Reddit) does not have a significant impact towards increasing contribution of users.

```
M3 = lm(posted ~ treated * first_timer, data = Q1)
summary(M3)
```

```
##
## Call:
## lm(formula = posted ~ treated * first_timer, data = Q1)
##
## 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
```

Here, the p-value for the interaction (0.0182) $<$ 0.05 (level of significance). This suggests that there is evidence to suggest that there is interaction between treated and first_timer, meaning that first_time users have a statistically significant effect on contribution. We also notice positive coefficient of interaction (0.108949), which means that first time users have a positive effect towards increasing contribution (more likely to increase their contribution) if provided with Reddit Gold.

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

The SUTVA (Stable Unit Treatment Value Assumption) assumption holds that there is no interference or spillover effect between treatment and control groups. In the aforementioned experiment, SUTVA can be

violated if:

1) Interference - If receiving Reddit Gold (treated = 1) generates observable influence on the behaviour of control group (for example, if a control group user, upon seeing increased attention received by Gold users modify their usage behaviour), then the assumption will be violated.

2) Spillover - If users who received Reddit Gold share their benefits/discuss the way they received Gold with control group (and/or) somehow change the overall atmoshphere of the platform, then SUTVA assumption will be violated.

3) Network/Community - If users use social network to obtain Gold advantages, resulting in influencing each other's behaviour and spread of Gold across the network, SUTVA assumption will be violated.

Q2) In 2019, Esther Duflo and Abhijeet Banerjee won the Nobel Prize in Economics for their research on experiments on education and poverty. In one of their experiments, they aimed to increase the academic performance of children in public schools in Vadodara (a town in India). Duflo and her co-authors examined the impact of the Balsakhi program. In the program, the weakest academic students in Grade 3 were pulled out of their classroom and provided with supplementary classes, during school hours, provided by a Balsakhi, a young woman from the community who would work with the children on basic skills. Schools that did not receive the program formed the comparison group.Data is provided for the period prior to the introduction of Balsakhis. This is known as the pre-period. Data is provided for math and language tests.

Using the data provided:

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).

```
Q2 = read_csv("data_Q2.csv")
```

```
## Rows: 91782 Columns: 53
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (11): div, sessiond, precheated, prehelped, prelater, midcheated, midhel...
## dbl (42): studentid, std, schoolid, divid, researchgroup, pre_tot, prepapers...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
std3 = subset(Q2, std == 3)
std3
```

```
## # A tibble: 46,206 x 53
##    studentid div     std schoolid divid researchgroup sessiond  precheated
##        <dbl> <chr> <dbl>    <dbl> <dbl>         <dbl> <chr>     <chr>
##  1     11095 B         3      108 31080             4 Morning   Yes
##  2     14729 A         3      120 31200             6 Afternoon No
##  3  40131243 A         3      401 34010             6 Afternoon No
##  4     13092 A         3      401 34010             6 Afternoon No
##  5     15316 B         3      117 31170             7 Afternoon Yes
##  6     14533 A         3      340 33400             7 Morning   No
##  7     12986 A         3      624 36240             3 Afternoon No
##  8     15318 B         3      117 31170             7 Afternoon No
##  9     16935 A         3      315 33150             2 Morning   No
## 10     10016 B         3      210 32100             5 Afternoon No
## # i 46,196 more rows
## # i 45 more variables: prehelped <chr>, prelater <chr>, pre_tot <dbl>,
## #   prepapersubtotal <dbl>, midcheated <chr>, midhelped <chr>, midlater <chr>,
## #   mid_tot <dbl>, midpapersubtotal <dbl>, postcheated <chr>, posthelped <chr>,
## #   postlater <chr>, post_tot <dbl>, postpapersubtotal <dbl>,
## #   attritprepost <dbl>, attritpremid <dbl>, attritmidpost <dbl>,
## #   pre_verb <dbl>, pre_math <dbl>, mid_verb <dbl>, mid_math <dbl>, ...
```

```
maths_pre = std3 %>% filter(test_type == 0, pre == 1)
t.test(norm ~ bal, data = maths_pre)
```

```
##
##  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
```

Here, we see that p-value for the test (0.3152) $>$ 0.05 (level of significance). Hence, we cannot reject null hypothesis, meaning that there is no statistical difference across pre-Balsakhi-period math test scores. There is no strong evidence to conclude statistically significant difference in the pre-period math test scores between treatment group and control group. Hence, we conclude that there is similarity in pre-Balsakhi-period math test scores of treatment and control groups.

```
language_pre = std3 %>%
  filter(test_type == 1, pre == 1)
t.test(norm ~ bal, data = language_pre)
```

```
##
##  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
```

Here, we see that p-value for the test (0.002466) < 0.05 (level of significance). Hence, we can reject null hypothesis, meaning that there is statistical difference between pre-Balsakhi-period language test scores across the two groups. There is evidence to conclude statistically significant difference in the pre-period language test scores between treatment group and control group.

Our analysis based on t-tests indicate that Balsakhi program had a significant impact on Language test scores but not Math scores during pre-period. This means that Balsakhi recepients had higher language test scores compared to non-recepients. Hence, we can conclude that randomization was not carried out effectively since there is a difference in scores of language tests between the treatment and control groups of students even prior to the introduction of Balsakhi program.

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.

```
maths_post <- std3 %>% filter(test_type == 0, post == 1)
t.test(norm ~ bal , maths_post)
```

```
##
##  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
```

Here, we see that p-value for the test (1.343e-07) < 0.05 (level of significance). Hence, we can reject null hypothesis, meaning that there is statistical difference between post-Balsakhi math test scores across the two groups. There is evidence to conclude statistically significant increase in the math scores of Balsakhi program recepients (treatment) compared to Non-Balsakhi recepients (control) for math tests.

```
language_post <- std3 %>%
  filter(test_type == 1, post == 1)
t.test(norm ~ bal , language_post)
```

```
##
##  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
```

Here, we see that p-value for the test (2.008e-05) < 0.05 (level of significance). Hence, we can reject null hypothesis, meaning that there is statistical difference between post-Balsakhi language test scores across the two groups. There is evidence to conclude statistically significant increase in the math scores of Balsakhi program recepients (treatment) compared to Non-Balsakhi recepients (control) for language tests.

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

For the pre-period, there was no significant statistical difference between treatment and control groups of math test, but a statistical difference between two groups for language test.

For the post-period, treatment and control groups of math and language tests both showed significant statistical difference amongst each other respectovely, as well as a significant improvement in scores post introduction of Balsakhi.

Results of the t-tests offer compelling support for the idea that the Balsakhi program had a statistically significant and favorable influence on both math and language examination scores in the post-period. Hence, we can conclude that the program resulted in an improvement in test scores for both subjects.

```
program <- std3 %>% filter(post==1)
M4 = lm(norm ~ bal, program)
summary(M4)
```

```
##
## Call:
## lm(formula = norm ~ bal, data = program)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8575 -0.9549 -0.2038  0.8231  3.4747
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.48743    0.01479  32.958   <2e-16 ***
## bal          0.17346    0.02088   8.307   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.176 on 12688 degrees of freedom
##   (2712 observations deleted due to missingness)
## Multiple R-squared:  0.005409,   Adjusted R-squared:  0.005331
## F-statistic: 69.01 on 1 and 12688 DF,  p-value: < 2.2e-16
```

Here, we see that p-value for the test (<2.2e-16) < 0.05 (level of significance). Hence, we can reject null hypothesis, meaning that there is statistical difference between post-Balsakhi test scores across the two groups. There is evidence to conclude statistically significant increase in the math and language scores of Balsakhi program recepients (treatment) compared to Non-Balsakhi recepients (control) for language tests.

d) Is the SUTVA assumption violated in the example?

The SUTVA (Stable Unit Treatment Value Assumption) assumption holds that there is no interference or spillover effect between treatment and control groups. In the aforementioned experiment, SUTVA can be violated if:

1) Interference - If recipients of Balsakhi interact with/generate observable influence on the performance of control group in tests (for example, sharing of knowledge learnt through the program), then the assumption will be violated.

2) Spillover - If recipients of Balsakhi somehow create spillover effect on control group (for example, changes in teaching methods for control group mimicking treatment group etc), then SUTVA assumption will be violated.

3) Network/Community - If behavior of recepients influencing behavior of control group (for example, control group motivated by treatment group's performance), SUTVA assumption will be violated.