Causal Assignment1

Ishan Kotian [5991337] and Swapnil Sharma[5955550]

2025-02-11

```
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
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.1     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
df_Q1 <- read.csv("/Users/ishankotian/Downloads/data_Q1.csv")
df_Q2 <- read.csv("/Users/ishankotian/Downloads/data_Q2.csv")
```

QUESTION 1 PART 1

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

```
summary(df_Q1 %>% select(tenure, premium_user, num_post_before))
```

```
##      tenure        premium_user    num_post_before
##  Min.   :   0.0   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:  23.0   1st Qu.:0.0000   1st Qu.: 0.000
##  Median : 253.5   Median :0.0000   Median : 0.000
##  Mean   : 549.9   Mean   :0.0221   Mean   : 1.597
##  3rd Qu.: 849.5   3rd Qu.:0.0000   3rd Qu.: 2.000
##  Max.   :4579.0   Max.   :1.0000   Max.   :45.000
```

To check if tenure is similar across treatment and control

```
t.test(df_Q1$tenure ~ df_Q1$treated)
```

```
##
##  Welch Two Sample t-test
##
## data:  df_Q1$tenure by df_Q1$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
```

The p-value is 0.1699, which is greater than 0.05. Since p-value is high, we fail to reject the null hypothesis. No statistical difference between treatment and control groups in terms of tenure.

To check if premium_user is similar across treatment and control

```
t.test(df_Q1$premium_user ~ df_Q1$treated)
```

```
##
##  Welch Two Sample t-test
##
## data:  df_Q1$premium_user by df_Q1$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
```

```
table(df_Q1$premium_user, df_Q1$treated)
```

```
##
##       0   1
##   0 882 888
##   1  23  17
```

The p-value is 0.3377, which is greater than 0.05. This means we fail to reject the null hypothesis, indicating that there is no statistically significant difference in the mean of premium_user between the treated and control groups.

To check if num_post_before is similar across treatment and control

```
t.test(df_Q1$num_post_before ~ df_Q1$treated)
```

```
##
##  Welch Two Sample t-test
##
## data:  df_Q1$num_post_before by df_Q1$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
```

The p-value is 0.5738, which is greater than 0.05, This means we fail to reject the null hypothesis,indicating that there is no statistically significant difference in the mean of num_post_before between the treated and control groups.

QUESTION 1 PART 2

Does getting reddit gold increase likelihood that the user will post

Fit a simple linear regression model

```
model <- lm(posted ~ treated, data = df_Q1)
summary(model)
```

Posted = beta0 + beta1[treated] + epsilon

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

Conclusions- The coefficient for treated is 0.06298 with a p-value of 0.0064, which is statistically significant at the 1% level ($p < 0.01$). This suggests that receiving Reddit Gold increases the likelihood of posting by approximately 6.3 percentage points on average.

```
model <- lm(posted ~ treated, data = df_Q1)
summary(model)
```

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

QUESTION 1 PART 3

What sorts of users are more likely to increase their contribution?

```
model_contribution <- lm(posted ~ tenure + first_timer, data = df_Q1)
summary(model_contribution)
```

Posted = beta0 + beta1[tenure] + beta2[first_timer] + epsilon

```
##
## Call:
## lm(formula = posted ~ tenure + first_timer, data = df_Q1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6650 -0.5723  0.3426  0.4274  0.5358
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.650e-01  2.108e-02  31.541  < 2e-16 ***
## tenure      -4.271e-05  1.717e-05  -2.488 0.012950 *
## first_timer -9.228e-02  2.378e-02  -3.880 0.000108 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4895 on 1807 degrees of freedom
## Multiple R-squared:  0.009659,   Adjusted R-squared:  0.008563
## F-statistic: 8.812 on 2 and 1807 DF,  p-value: 0.0001555
```

Key Findings: 1. Intercept (Baseline Posting Probability): The intercept is 0.665, which indicates that the likelihood of posting is 66.5% when tenure = 0 and first_timer = 0 (a non-first-time user without tenure). Since both cannot be equal to zero, this cannot be interpreted.

2. The Impact of Tenure: The coefficient is -4.271e-05, which is statistically significant at the 5% level (p-value = 0.01295). Longer tenure may somewhat lower the chance of posting, according to a negative coefficient. However, tenure has little practical impact because of the extremely modest effect size.

3. Impact of First-Time User Status (first_timer): Coefficient: -0.09228 (p-value = 0.000108, 0.1% level of statistical significance). This suggests that compared to returning users, first-time users are 9.23 percentage points less likely to post.

QUESTION 1 PART 4

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

Positive Influence In this case, where Reddit Gold users are the treatment group and non-users are the control group, interference occurs when a member of the control group is influenced by viewing the Reddit Gold status of others. For instance, even though they are in the control group, the treatment (Reddit Gold status) has affected their behaviour if a control group member chooses to post or remark more frequently than usual after viewing a post from a Reddit Gold user.

Since this illustrates peer influence or the spillover effect, it immediately violates the SUTVA's no interference rule. Therefore, it is likely that the experiment will violate the SUTVA assumption.

QUESTION 2 PART 1

For Mathematics-

```r
pre_math <- subset(df_Q2, pre == 1 & test_type == 0 & std == 3)
t_test_math <- t.test(norm ~ bal, data = pre_math)
print(t_test_math)
```

```
##
##  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
```

Since the p-value is much greater than 0.05 (0.3152), we fail to reject the null hypothesis. This indicates that the pre-period math scores of the treatment and control groups do not differ statistically significantly.The Means Are Almost the Same Students in the treatment and control groups had comparable baseline math scores, as evidenced by the two groups' mean scores being quite near to zero.

Randomization Appears Valid

For Language-

```
pre_lang <- subset(df_Q2, pre == 1 & test_type == 1 & std == 3)
t_test_lang <- t.test(norm ~ bal, data = pre_lang)
print(t_test_lang)
```

```
##
##  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
```

Null Hypothesis :There is no significant difference in the average normalized language test scores between the treatment group.

Alternative Hypothesis: There is a significant difference in the average normalized language test scores between the treatment and control groups in the pre-period.

Conclusion: Since the p-value (0.002466) is less than 0.05, we reject the null hypothesis. This means there is a statistically significant difference in the pre-period language test scores between the treatment and control groups.

Implication for Randomization Check: The existence of a statistically significant discrepancy raises the possibility that the randomisation process was not flawless. Prior to the implementation of the Balsakhi program, test results for the treatment and control groups should ideally not differ significantly. The observed difference, however, suggests that the baseline language scores of the pupils in the treatment and control groups could not have been quite equal. The validity of subsequent effect assessments of the program may be impacted by this.

QUESTION 2 PART 2

For Mathematics

```
post_math <- subset(df_Q2, post == 1 & test_type == 0 & std == 3)
t_test_post_math <- t.test(norm ~ bal, data = post_math)
print(t_test_post_math)
```

```
##
##  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
```

Null Hypothesis : There is no significant difference in post-period math scores between treatment and control groups. Alternative Hypothesis: There is a significant difference in post-period math scores between treatment and control groups.

The p-value is extremely small (less than 0.05), meaning we reject the null hypothesis. This indicates a statistically significant difference in math scores between the two groups. The treatment group (bal = 1) has a higher average score (0.4341) compared to the control group (0.2535).

Conclusion: The Balsakhi program significantly improved math performance in the post-period.

For Language

```
post_lang <- subset(df_Q2, post == 1 & test_type == 1 & std == 3)
t_test_post_lang <- t.test(norm ~ bal, data = post_lang)
print(t_test_post_lang)
```

```
##
##  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
```

Null Hypothesis: There is no significant difference in post-period language scores between treatment and control groups. Alternative Hypothesis: There is a significant difference in post-period language scores between treatment and control groups.

The p-value is extremely small (less than 0.05), meaning we reject the null hypothesis. This indicates a statistically significant difference in language scores between the two groups. The treatment group (bal = 1) has a higher average score (0.8742) compared to the control group (0.7152).

Conclusion: The Balsakhi program significantly improved language performance in the post-period, similar to the results observed for math.

Post-Period (Impact Analysis): The treatment group's scores on the language and math tests were noticeably higher than those of the control group. Both subjects experienced positive improvements in student performance as a result of the Balsakhi program.

QUESTION 2 PART 3

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

The t-tests confirm that students in the treatment group performed significantly better than those in the control group in both math and language after the program.

But in order to prove causation, we need to make sure that the program alone was responsible for the score increases and not other influences. The Balsakhi program seems to be helping students, at least on the surface. It's crucial to remember that there was a small variation between the groups in the pre-period as well, which might have affected the outcomes.

For Math

```
did_math <- lm(norm ~ bal * post, data = subset(df_Q2, test_type == 0))
summary(did_math)
```

```
##
## Call:
## lm(formula = norm ~ bal * post, data = subset(df_Q2, test_type ==
##     0))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3011 -0.9123 -0.0598  0.8345 11.3494
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.18370    0.01110  16.548  < 2e-16 ***
## bal          0.05740    0.01584   3.624 0.000291 ***
## post        -0.01308    0.02002  -0.653 0.513534
## bal:post     0.11955    0.02870   4.165 3.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.098 on 27656 degrees of freedom
##   (2934 observations deleted due to missingness)
## Multiple R-squared:  0.002791,   Adjusted R-squared:  0.002682
## F-statistic:  25.8 on 3 and 27656 DF,  p-value: < 2.2e-16
```

For Language

```
did_lang <- lm(norm ~ bal * post, data = subset(df_Q2, test_type == 1))
summary(did_lang)
```

```
##
## Call:
## lm(formula = norm ~ bal * post, data = subset(df_Q2, test_type ==
##     1))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4354 -0.9582 -0.2241  0.8939  8.2123
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.35276    0.01199  29.414  < 2e-16 ***
## bal          0.06840    0.01709   4.001 6.31e-05 ***
```

```
## post         0.31401    0.02160  14.535  < 2e-16 ***
## bal:post     0.05886    0.03096   1.901   0.0573 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.184 on 27620 degrees of freedom
##   (2970 observations deleted due to missingness)
## Multiple R-squared:  0.01879,    Adjusted R-squared:  0.01868
## F-statistic: 176.3 on 3 and 27620 DF,  p-value: < 2.2e-16
```

Conclusion: Math Scores: Significant improvement caused by the Balsakhi program (p = 3.12e-05). Language Scores: Weak evidence of an effect (p = 0.0573, slightly above 0.05).

QUESTION 2 PART 4

Is the SUTVA assumption violated in the example?

The SUTVA makes the assumption that there is no difference in therapy between treated units and that there is no interference between units, which means that the outcome of one student's treatment shouldn't affect that of another. This presumption might be broken in the Balsakhi program, though, because of possible behavioural changes and spillover consequences. Teachers may offer more assistance to difficult students who are not enrolled in the program if they are aware that some kids are getting more support. The no-interference presumption is violated by this indirect benefit to kids who were not treated.

Furthermore, the program's existence can cause teachers and students to behave differently. Beyond the intervention's immediate effects, academic performance may be impacted by teachers changing their teaching strategies or by students outside the treatment group changing their learning habits. Because of this, the estimated treatment impact may be skewed—it may be exaggerated if treated students obtained benefits outside of the program, or it may be understated if non-treated students benefited from spillover effects.

Hence, SUTVA assumption is likely to be voilated.