# DIL-Data Task

Yu Hui

2024-06-22

## Load Data

```
df_outcome <- read.csv("student_follow_ups.csv")
df_stu_base<-read.csv("student_baseline.csv")
df_sch_base<-read.csv("schools.csv")
df_visit<-read.csv("school_visits_log.csv")
```

## Detect Missing Values

```
missing_percentages <- df_outcome %>%
  summarise_all(~ mean(is.na(.)) * 100)
missing_percentages <- as.data.frame(t(missing_percentages))
colnames(missing_percentages) <- "Missing_Percentage"
print(missing_percentages)
```

```
##            Missing_Percentage
## school_id          0.00000000
## student_id         0.00000000
## year               0.00000000
## died               0.02063771
## married            0.02063771
## children          49.57176762
## pregnant          49.57176762
## dropout            0.37147869
```

It seems like outcome of pregnancy and whether have children have great amount of missing values, thus these two outcome variable may be carefully considered when processed regression.

## Concatenate Datasets

```
merged_df <- merge(df_outcome, df_stu_base, by = "student_id", all.x = TRUE)
merged_df <- merge(merged_df,df_sch_base,by="school_id",all.x=TRUE)
df_visit<- df_visit %>% select(-year)
merged_df <- merge(merged_df,df_visit,by="school_id",all.x=TRUE)
```

## Balance Table

```r
### school level covariates
df_sch_base_1 <-df_sch_base[,-c(1:3)] %>%
  mutate(Urban = ifelse(location == 1, 1, 0))%>%
  select(-location)
table <-df_sch_base_1 %>%
  group_by(treatment)%>%
  summarize_all(list(~mean(.))) %>%
  mutate_all(list(~round(., 3))) %>%
  t() %>%
  as.data.frame()
colnames(table) <- c("mean_control", "mean_treat")

p_values <- data.frame(variable = rownames(table), p_value = NA)

# Calculate p-values using t-tests
for (var in rownames(table)[-1]) { # Exclude the treatment row itself
  control_values <- df_sch_base_1 %>% filter(treatment == 0) %>% pull(var)
  treat_values <- df_sch_base_1 %>% filter(treatment == 1) %>% pull(var)

  # Perform t-test
  t_test_result <- t.test(control_values, treat_values)

  # Store p-value
  p_values$p_value[p_values$variable == var] <- t_test_result$p.value
}

# Merge the means and p-values
final_table_A <- cbind(table, p_values[,-1])
final_table_A<-final_table_A[-1,]

# Display the final table
knitr::kable(final_table_A)
```

| | mean_control | mean_treat | p_values[, -1] |
|---|---|---|---|
| n_teachers | 14.537 | 14.181 | 0.5788467 |
| n_teachers_fem | 7.146 | 7.205 | 0.9259432 |
| female_head_teacher | 0.085 | 0.096 | 0.8069064 |
| n_students_fem | 254.963 | 231.675 | 0.1283670 |
| n_students_male | 253.573 | 232.952 | 0.1713085 |
| n_schools_2km | 1.951 | 2.012 | 0.8315288 |
| av_teacher_age | 37.922 | 38.287 | 0.8808257 |
| av_student_score | 245.024 | 242.422 | 0.7849296 |
| n_latrines | 5.585 | 11.627 | 0.0314354 |
| Urban | 0.098 | 0.157 | 0.2571961 |

The balance table shows the means of covariates for the control and treatment groups.P-values indicate if the difference in means between groups is statistically significant.In this table, only n_latrines shows a significant difference (p = 0.0314354), suggesting a potential imbalance in the number of latrines between the groups.The balance table indicates that the randomization was generally successful, with the exception of n_latrines. This imbalance should be addressed in the analysis phase to ensure it does not confound the treatment effects.

```r
## indivial level covariates
table_ind <- merged_df %>%
  filter(year==3) %>%
```

```r
  select(c(sex,yob,treatment)) %>%
  mutate(sex = ifelse(sex == 1, 1, 0))%>%
  group_by(treatment)%>%
  summarize_all(list(~mean(.))) %>%
  mutate_all(list(~round(., 3))) %>%
  t() %>%
  as.data.frame()
colnames(table_ind) <- c("mean_control", "mean_treat")

# Function to calculate p-values
calc_p_value <- function(var) {
  control_values <- merged_df %>% filter(year == 3 & treatment == 0) %>% pull(!!sym(var))
  treat_values <- merged_df %>% filter(year == 3 & treatment == 1) %>% pull(!!sym(var))
  t.test(control_values, treat_values)$p.value
}

# Calculate p-values for the relevant columns
variables <- c("sex", "yob")
p_values <- map_dbl(variables, calc_p_value)

# Create a dataframe for p-values
p_value_df <- data.frame(variable = rownames(table_ind)[-1], p_value = p_values)
rownames(p_value_df) <- p_value_df$variable
p_value_df$variable <- NULL

# Merge the p-values with the table of means
final_table_B <- cbind(table_ind, p_value = c(NA, p_values))
final_table_B<-final_table_B[-1, ]


# Display the final table
knitr::kable(final_table_B)
```

| | mean_control | mean_treat | p_value |
|---|---|---|---|
| sex | 0.502 | 0.491 | 0.1350085 |
| yob | 2563.124 | 2520.784 | 0.1463867 |

The balance table indicates that the randomization was generally successful across individual level confounders.

## Regression

We will choose cross-section linear regression model to study the impact of the trearment.The unit of analysis is individual students. The regresssion specfication is:

$$\text{Outcome}_{it} = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 \text{BaselineCovariates}_i + \gamma_d + \epsilon_{it} \tag{1}$$

Where:

$\text{Outcome}_{it}$: The outcome of interest (school evasion, teen pregnancy, marriage) for student $i$ at time $t$ (3 or 5 years).

$\text{Treatment}_i$: Treatment indicator for student $i$ (1 if treated, 0 if control).

$selineCovariates_i$: Baseline covariates controls for student $i$ (The control variables include school level controls and student level controls. Specifically, they are sex, age as individual characteristics, number of teachers, number of female teachers, number of female head teachers, number of female students, number of school within 2km, average teacher age, average student score, number of latrines as school level characteristics. To avoid muticonlinearity between total number of teachers, number of female teachers and number of male teachers, exculde number of male teachers in control selection).

$\gamma_d$: District fixed effects.(urban, district, stratum and month fixed effect varaibles )

$\epsilon_{it}$: Error term.

```
merged_df_3_example<- merged_df %>%
  filter(year == 3)
m <- merged_df_3_example %>%
  group_by(died) %>%
  summarize_all(list(~mean(.))) %>%
  mutate_all(list(~round(., 3))) %>%
  t()
colnames(m) <-c("Unknown","Alive","Died","NA")
knitr::kable(m)
```

| | Unknown | Alive | Died | NA |
|---|---|---|---|---|
| died | -99.000 | 0.000 | 1.000 | NA |
| school_id | 72.689 | 80.705 | 58.963 | 81.000 |
| student_id | 13538984.975 | 15896858.715 | 5956328.852 | 8160055.000 |
| year | 3.000 | 3.000 | 3.000 | 3.000 |
| married | -59.823 | -0.140 | -99.000 | NA |
| children | NA | NA | NA | NA |
| pregnant | NA | NA | NA | NA |
| dropout | 0.592 | 0.146 | NA | NA |
| sex | 1.589 | 1.499 | 1.296 | 1.500 |
| yob | 5213.415 | 2371.286 | 4659.444 | 5995.000 |
| district | 1.488 | 1.592 | 1.444 | 1.500 |
| stratum | 50.043 | 43.888 | 50.519 | 55.500 |
| treatment | 0.499 | 0.491 | 0.444 | 0.500 |
| location | 1.800 | 1.844 | 1.778 | 2.000 |
| n_teachers | 16.440 | 15.521 | 14.926 | 12.500 |
| n_teachers_fem | 8.996 | 7.962 | 8.667 | 7.000 |
| female_head_teacher | 0.066 | 0.068 | 0.037 | 0.500 |
| n_students_fem | 293.395 | 279.388 | 253.963 | 246.500 |
| n_students_male | 296.610 | 279.556 | 254.593 | 263.500 |
| n_schools_2km | 1.606 | 1.806 | 1.296 | 2.500 |
| av_teacher_age | 38.436 | 38.515 | 34.819 | 36.874 |
| av_student_score | 251.862 | 252.520 | 223.963 | 217.000 |
| n_latrines | 11.242 | 10.863 | 5.741 | 13.000 |
| day | 20.169 | 19.819 | 20.741 | 21.500 |
| month | 7.589 | 7.728 | 7.574 | 7.500 |

We can see that all died student have no/unknown data of outcome variables including marriage and drop out, thus removing the died student will not cause much loss in outcome data, we can safely remove them.

```
merged_df_3<- merged_df %>%
  filter(year == 3 ,died == 0) %>%
  mutate(age = 2010-yob) %>% ## create col of age instead of yob, which is more interpretable
  filter(age > 0) %>%
```

```
  mutate(urban = ifelse(location == 1, 1, 0))%>%
  mutate(sex = ifelse(location == 1, 1, 0))%>%
  mutate(stratum = as.factor(stratum)) %>%
  mutate(district = as.factor(district)) %>%
  mutate(district = as.factor(month)) %>%
  select(-c(school_id,student_id,year,yob,location))
```

```
##(merged_df_3)
```

We can see that there is no

```
model.3.mar <- lm(married ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + n_st
model.3.pregnant <- lm(pregnant ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher
model.3.children <- lm(children ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher


#summary(model.3.dropout)
stargazer(model.3.mar,model.3.pregnant,model.3.children, type = "text",
            omit = "stratum", omit.labels = "stratum",
            star.cutoffs = c(0.05, 0.01, 0.001),
            #dep.var.labels = c("Outcome"),
            keep.stat = c("n", "rsq", "adj.rsq", "f"),
            single.row = TRUE,
            title = "Regression Results -- 3 year")
```

```
##
## Regression Results -- 3 year
## ================================================================================================
##                                              Dependent variable:
##                      ------------------------------------------------------------------------
##                                married                  pregnant                  children
##                                  (1)                      (2)                       (3)
## ------------------------------------------------------------------------------------------------
## treatment                0.158* (0.077)           0.275 (0.166)             0.241 (0.124)
## sex                       0.118 (0.185)          1.140** (0.388)            0.736* (0.290)
## age                     -0.079** (0.025)          -0.112 (0.059)            -0.017 (0.044)
## n_teachers               -0.008 (0.029)          -0.195** (0.064)          -0.105* (0.047)
## n_teachers_fem           -0.045 (0.025)           0.054 (0.054)             0.032 (0.040)
## female_head_teacher      0.481* (0.192)           -0.617 (0.409)            0.329 (0.305)
## n_students_fem           -0.0005 (0.001)          0.003 (0.002)             0.001 (0.001)
## n_schools_2km             0.011 (0.034)           0.078 (0.075)             0.090 (0.056)
## av_teacher_age            0.003 (0.004)           0.0004 (0.008)            0.005 (0.006)
## av_student_score         0.0002 (0.001)           -0.004 (0.002)            0.003 (0.002)
## n_latrines              -0.009** (0.004)         0.026*** (0.008)          -0.005 (0.006)
## urban
## district5                 0.054 (0.193)           0.145 (0.410)             0.015 (0.306)
## district7                 0.125 (0.215)           -0.579 (0.451)            -0.229 (0.337)
## district9                 0.035 (0.152)           0.092 (0.322)             0.015 (0.240)
## district10               -0.031 (0.205)           -0.495 (0.433)            -0.365 (0.323)
## district11                0.210 (0.231)           -0.469 (0.486)            -0.116 (0.363)
## district12               -0.107 (0.397)          -2.381** (0.876)           -0.873 (0.654)
## month
## Constant                  1.009 (0.840)          6.169*** (1.832)           -0.483 (1.367)
## ------------------------------------------------------------------------------------------------
## stratum                       Yes                      Yes                       Yes
```

```
## ------------------------------------------------------------------------------------------------
## Observations                17,342                    8,532                    8,532
## R2                           0.016                    0.030                    0.019
## Adjusted R2                  0.010                    0.019                    0.008
## F Statistic         2.827*** (df = 98; 17243) 2.666*** (df = 98; 8433) 1.684*** (df = 98; 8433)
## ================================================================================================
## Note:                                                              *p<0.05; **p<0.01; ***p<0.001
```

```
model.3.dropout <- lm(dropout ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + n
stargazer(model.3.dropout, type = "text",
            omit = "stratum", omit.labels = "stratum",
            star.cutoffs = c(0.05, 0.01, 0.001),
            #dep.var.labels = c("Outcome"),
            keep.stat = c("n", "rsq", "adj.rsq", "f"),
            single.row = TRUE,
            title = "Regression Results -- 3 year")
```

```
##
## Regression Results -- 3 year
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                dropout
## ------------------------------------------------
## treatment                -0.016** (0.005)
## sex                       -0.021 (0.013)
## age                      0.069*** (0.002)
## n_teachers                 0.002 (0.002)
## n_teachers_fem             0.002 (0.002)
## female_head_teacher       0.035** (0.013)
## n_students_fem           0.00001 (0.0001)
## n_schools_2km              0.004 (0.002)
## av_teacher_age           -0.0002 (0.0003)
## av_student_score        -0.0002** (0.0001)
## n_latrines               -0.001* (0.0003)
## urban
## district5                 -0.002 (0.013)
## district7                  0.024 (0.015)
## district9                 -0.002 (0.010)
## district10                 0.006 (0.014)
## district11                0.034* (0.016)
## district12                 0.013 (0.027)
## month
## Constant                 -1.320*** (0.058)
## ------------------------------------------------
## stratum                          Yes
## ------------------------------------------------
## Observations                    17,342
## R2                               0.110
## Adjusted R2                      0.105
## F Statistic          21.755*** (df = 98; 17243)
## ================================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

For three year sample, as we can see from the table above, the coefficient before treatment is not significant (at 5% level) for outcome variable: marriage, children and pregnant, implying that the intervention did not change these

concerns evidently. However, we can see a significant negative impact of treatment on drop-out rate, meaning that offer transfers can decrease students' dropout.

```
## effect after 5 years
merged_df_5<- merged_df %>%
  filter(year == 5 ,died == 0) %>%
  mutate(age = 2010-yob) %>% ## create col of age instead of yob, which is more interpretable
  filter(age > 0) %>%
  mutate(urban = ifelse(location == 1, 1, 0))%>%
  mutate(sex = ifelse(location == 1, 1, 0))%>%
  mutate(stratum = as.factor(stratum)) %>%
  mutate(district = as.factor(district)) %>%
  mutate(district = as.factor(month)) %>%
  select(-c(school_id,student_id,year,yob,location))
```

```
model.5.mar <- lm(married ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + n_st

model.5.pregnant <- lm(pregnant ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher

model.5.children <- lm(children ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher

model.5.dropout <- lm(dropout ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + 
#summary(model.3.dropout)
stargazer(model.5.mar,model.5.pregnant,model.5.children,model.5.dropout, type = "text",
            omit = "stratum", omit.labels = "stratum",
            star.cutoffs = c(0.05, 0.01, 0.001),
            #dep.var.labels = c("Outcome"),
            keep.stat = c("n", "rsq", "adj.rsq", "f"),
            single.row = TRUE,
            title = "Regression Results Excluding Stratum Coefficients")
```

```
##
## Regression Results Excluding Stratum Coefficients
## ================================================================================================
##                                                  Dependent variable:
##                     ---------------------------------------------------------------------------
##                             married                 pregnant                children
##                               (1)                     (2)                     (3)
## ------------------------------------------------------------------------------------------------
## treatment                -0.133 (0.138)          -0.371 (0.303)          0.266 (0.274)          -0
## sex                       0.270 (0.342)          0.720 (0.724)           0.337 (0.654)          -0
## age                    -0.172*** (0.044)       -0.232* (0.107)         -0.194* (0.097)          0.
## n_teachers                0.100 (0.053)          0.125 (0.116)           0.105 (0.105)           0
## n_teachers_fem           -0.084 (0.046)          -0.072 (0.100)          -0.003 (0.090)         -
## female_head_teacher      -0.054 (0.342)          -0.383 (0.731)          0.693 (0.661)           0.
## n_students_fem           -0.004* (0.002)       -0.009** (0.004)        -0.007* (0.003)          0.0
## n_schools_2km             0.118 (0.065)          -0.002 (0.146)          0.215 (0.132)           0.
## av_teacher_age            0.005 (0.007)          0.0002 (0.016)          0.007 (0.015)          -0
## av_student_score          0.002 (0.002)          -0.007 (0.004)          0.001 (0.004)         -0.0
## n_latrines                0.001 (0.006)          -0.004 (0.014)          -0.010 (0.012)         -0.
## urban
## district5                 0.295 (0.348)          0.418 (0.746)           0.195 (0.674)           0
## district7                -0.194 (0.391)          -0.438 (0.829)          0.084 (0.750)           0
## district9                 0.198 (0.276)          0.290 (0.589)           0.054 (0.533)           0
## district10               -0.243 (0.387)          -0.251 (0.821)          -0.055 (0.743)          0
## district11               -0.147 (0.420)          -0.344 (0.892)          0.428 (0.807)           0
```

```
## district12               -0.551 (0.719)           -2.860 (1.632)           -1.516 (1.476)            0
## month
## Constant                  2.706 (1.494)           6.698* (3.266)           1.584 (2.953)           -1.
## ------------------------------------------------------------------------------------------------------
## stratum                        Yes                      Yes                      Yes
## ------------------------------------------------------------------------------------------------------
## Observations                15,734                   7,708                    7,708
## R2                           0.018                    0.025                    0.024
## Adjusted R2                  0.012                    0.013                    0.011
## F Statistic         2.896*** (df = 98; 15635) 2.008*** (df = 98; 7609) 1.874*** (df = 98; 7609) 24.280*
## ======================================================================================================
## Note:                                                                                  *p<0.05; *
```

```
model.5.dropout <- lm(dropout ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + 
stargazer(model.5.dropout, type = "text",
            omit = "stratum", omit.labels = "stratum",
            star.cutoffs = c(0.05, 0.01, 0.001),
            #dep.var.labels = c("Outcome"),
            keep.stat = c("n", "rsq", "adj.rsq", "f"),
            single.row = TRUE,
            title = "Regression Results -- 3 year")
```

```
##
## Regression Results -- 3 year
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                dropout
## ------------------------------------------------
## treatment                -0.020** (0.007)
## sex                      -0.046** (0.017)
## age                      0.087*** (0.002)
## n_teachers                 0.001 (0.003)
## n_teachers_fem            -0.001 (0.002)
## female_head_teacher      0.073*** (0.017)
## n_students_fem           0.0003*** (0.0001)
## n_schools_2km              0.009** (0.003)
## av_teacher_age           -0.0004 (0.0003)
## av_student_score        -0.0004*** (0.0001)
## n_latrines               -0.001** (0.0003)
## urban
## district5                  0.020 (0.017)
## district7                  0.030 (0.019)
## district9                  0.011 (0.013)
## district10                 0.019 (0.019)
## district11                 0.032 (0.020)
## district12                 0.068 (0.035)
## month
## Constant                 -1.727*** (0.072)
## ------------------------------------------------
## stratum                          Yes
## ------------------------------------------------
## Observations                  15,734
## R2                             0.132
## Adjusted R2                    0.127
## F Statistic           24.280*** (df = 98; 15635)
```

```
## ================================================
## Note:                *p<0.05; **p<0.01; ***p<0.001
```

For five year sample, as we can see from the table above, the coefficient before treatment is not significant (at 5% level) for outcome variable: marriage, children and pregnant, implying that the intervention did not change these concerns evidently. However, we can still see a significant impact of treatment on drop-out rate. Moreover, the impact is greater than 3 year situation, meaning that the effects of treatment amplifed overtime.

## Further Analysis

I want to seperately analyze the impact of treatment on male and female students.

```
merged_df_5_f<- merged_df_5 %>% filter(sex == 0)
model.5.dropout <- lm(dropout ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + n
stargazer(model.5.dropout, type = "text",
            omit = "stratum", omit.labels = "stratum",
            star.cutoffs = c(0.05, 0.01, 0.001),
            #dep.var.labels = c("Outcome"),
            keep.stat = c("n", "rsq", "adj.rsq", "f"),
            single.row = TRUE,
            title = "Regression Results -- 3 year")
```

```
##
## Regression Results -- 3 year
## ================================================
##                            Dependent variable:
##                        ----------------------------
##                                    dropout
## ------------------------------------------------
## treatment                    -0.005 (0.008)
## sex
## age                         0.089*** (0.002)
## n_teachers                   0.0001 (0.003)
## n_teachers_fem               -0.002 (0.002)
## female_head_teacher          0.044* (0.019)
## n_students_fem              0.001*** (0.0001)
## n_schools_2km               0.013*** (0.003)
## av_teacher_age              -0.0004 (0.0005)
## av_student_score           -0.001*** (0.0001)
## n_latrines                  -0.001* (0.0003)
## urban
## district5                     0.021 (0.020)
## district7                     0.022 (0.022)
## district9                     0.014 (0.016)
## district10                    0.003 (0.020)
## district11                    0.029 (0.023)
## district12                    0.018 (0.038)
## month
## Constant                    -1.763*** (0.077)
## ------------------------------------------------
## stratum                           Yes
## ------------------------------------------------
## Observations                    13,480
## R2                               0.132
## Adjusted R2                      0.126
```

```
## F Statistic          21.706*** (df = 94; 13385)
## ================================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

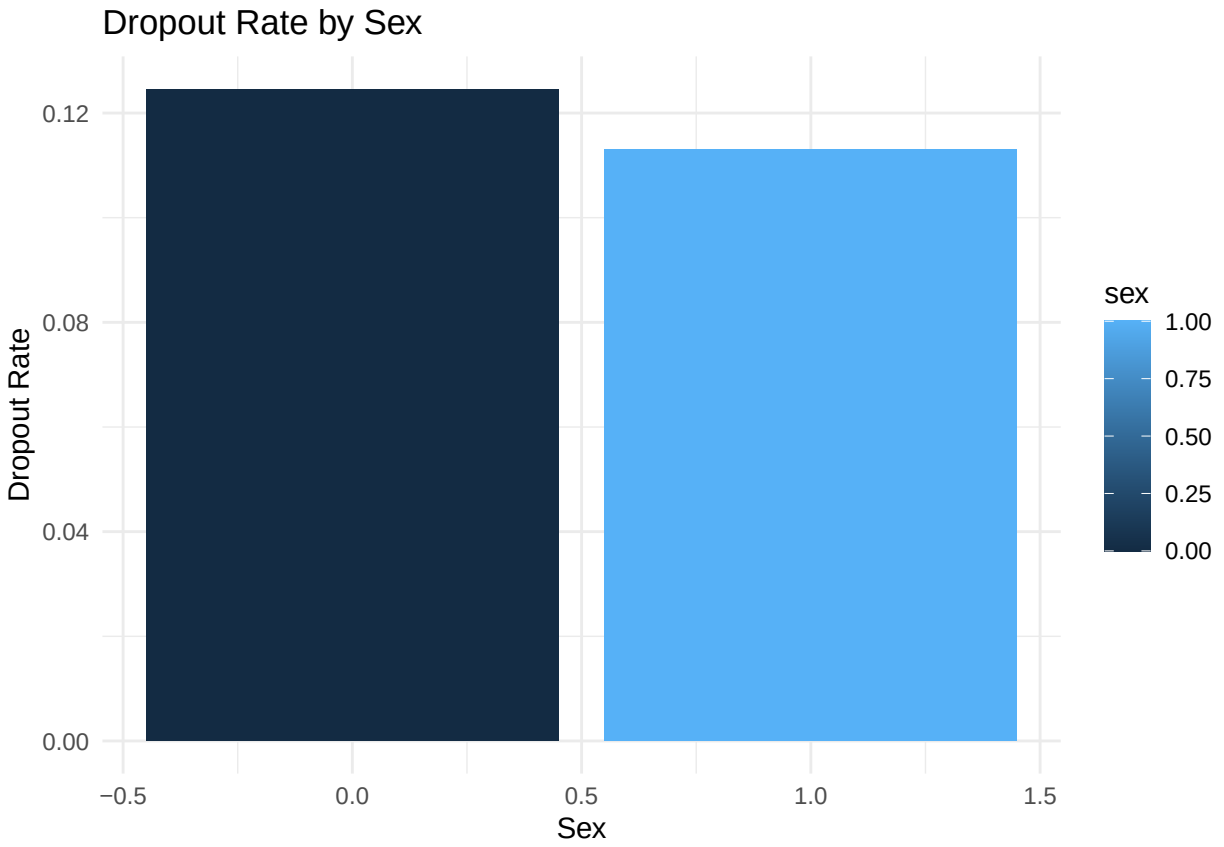
```
merged_df_5_m<- merged_df_5 %>% filter(sex == 1)
model.5.dropout <- lm(dropout ~ treatment+ sex + age + n_teachers + n_teachers_fem+ female_head_teacher + 
stargazer(model.5.dropout, type = "text",
            omit = "stratum", omit.labels = "stratum",
            star.cutoffs = c(0.05, 0.01, 0.001),
            #dep.var.labels = c("Outcome"),
            keep.stat = c("n", "rsq", "adj.rsq", "f"),
            single.row = TRUE,
            title = "Regression Results -- 3 year")
```

```
##
## Regression Results -- 3 year
## ================================================
##                           Dependent variable:
##                       -----------------------------
##                                  dropout
## ------------------------------------------------
## treatment                 -1.180*** (0.294)
## sex
## age                       0.081*** (0.005)
## n_teachers                -0.422*** (0.119)
## n_teachers_fem            0.033*** (0.009)
## female_head_teacher       1.109*** (0.285)
## n_students_fem             0.004** (0.001)
## n_schools_2km             -1.005*** (0.271)
## av_teacher_age            -0.007** (0.002)
## av_student_score          -0.010** (0.003)
## n_latrines                0.215*** (0.052)
## urban
## district5                  -0.000 (0.040)
## district7                  -0.000 (0.041)
## district9                  -0.000 (0.024)
## district10                 -0.000 (0.063)
## district11                 -0.000 (0.051)
## month
## Constant                   8.408** (2.897)
## ------------------------------------------------
## stratum                        Yes
## ------------------------------------------------
## Observations                  2,254
## R2                            0.163
## Adjusted R2                   0.154
## F Statistic           16.733*** (df = 26; 2227)
## ================================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

```
library(ggplot2)
dropout_rate_by_sex <- merged_df_3 %>%
  group_by(sex) %>%
  summarize(dropout_rate = mean(dropout, na.rm = TRUE))

# Create the plot
```

```r
ggplot(dropout_rate_by_sex, aes(x = sex, y = dropout_rate, fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Dropout Rate by Sex",
       x = "Sex",
       y = "Dropout Rate") +
  theme_minimal()
```




(should turn sex into categorical variable but run out of time..)

As we can see from the analysis above, the impact of treatment mostly come from change in dropout rate of male students, not female. and from the graph above, we can see that for male student have lower drop out rate.

Further, I want to conduct the same analysis on variable including married, pregnant and children, to see if there is also hetergenous treatment effects.