Homework 1

Setup

All questions below are based on the paper “Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment,” by Karlan and List, *The American Economic Review* (2007).

Please complete the code-chunk sections as well as written answers in this document. When you finalized the code and your answers, compile the markdown file into a PDF document and submit the PDF via CCLE. If necessary, you may compile from markdown to Word, and then “print to PDF” from the Word document.

1. Table 1

1.1

Load the “charitable_giving.csv” dataset and run a regression to assess whether the average “Number of months since last donation” is significantly different between treatment and control. Interpret the relevant regression coefficients and compare the regression-based comparison to the group-specific means reported in Table 1 of the paper.

```
df = read.csv("charitable_giving.csv", header = TRUE)
reg_1_1 = lm(months_since_last_donation~treatment, data = df)
summary(reg_1_1)

##
## Call:
## lm(formula = months_since_last_donation ~ treatment, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.012  -9.012  -5.012   6.002 154.988
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.99814    0.09353 138.979   <2e-16 ***
## treatment    0.01369    0.11453   0.119    0.905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.08 on 50080 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  2.851e-07,  Adjusted R-squared:  -1.968e-05
## F-statistic: 0.01428 on 1 and 50080 DF,  p-value: 0.9049
```

The intercept of linear regression represents the mean of "months since last donation" of control group, which is 12.99814 months; the intercept is the original condition when people don't receive the treatment. The mean for the treatment group is 12.99814 + 0.01369 = 13.01183 months. We don't need to run two regression of both treatment and control since one already implies the comparison.

Looking at the means of both groups in the table, they are indeed the same as the what I got from the regression.

1.2

Is the difference in "Number of month since last donation" between treatment and control statistically significant (at the usual 95% confidence level)? Is this the result you expected?

No, it's not. The p value is 0.905 so at 95% significance level 0.905>0.05, which means treatment isn't statistically significant to affect the average number of months since the last donation.

This is the result I expected since the treatment and control groups were randomly assigned. No statistically significant difference should be a consequence.

1.3

More generally, describe the take-away from Table 1 in the paper.

It's an imbalanced dataset. To decide if the treatment group is statistically significant from the control group, just using one variable isn't enough; other variables in the table also need to be taken into account. A general observation is that randomization ensured that there are not statistically significant differences between the demographics represented in the treatment and control groups (which may requires further confirmation).

2. Response rate regressions

2.1

Run a linear regression of response rate (the donation dummy) on the treatment dummy (and an intercept). Interpret both coefficients and compare them to the results presented in the first row of Table 2a.

```
reg_2_1 = lm(donation_dummy~treatment, data = df)
summary(reg_2_1)

##
## Call:
## lm(formula = donation_dummy ~ treatment, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.02204 -0.02204 -0.02204 -0.01786  0.98214 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 0.017858   0.001101  16.225  < 2e-16 ***
## treatment   0.004180   0.001348   3.101  0.00193 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.1422 on 50081 degrees of freedom
##   (47 observations deleted due to missingness)
## Multiple R-squared:  0.000192,   Adjusted R-squared:  0.0001721 
## F-statistic: 9.618 on 1 and 50081 DF,  p-value: 0.001927
```

The intercept of linear regression represents the mean of "response rate" of control group, which is 0.017858; the intercept is the original condition when people don't receive the treatment. The mean for the treatment group is $0.017858 + 0.004180 = 0.022038$ months. They are consistent as the corresponding mean in table 2a. At a significance level of 95%, the p value of treatment is less than 0.05, which means it's statistically significant.

2.2

Run a regression on three dummies for match ratio treatment (1:1, 2:1, and 3:1 and an intercept). Interpret all four regression coefficients.

```
reg_2_2 = lm(donation_dummy~ratio1+ratio2+ratio3, data = df)
summary(reg_2_2)
```

```
## 
## Call:
## lm(formula = donation_dummy ~ ratio1 + ratio2 + ratio3, data = df)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.02273 -0.02263 -0.02075 -0.01786  0.98214 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 0.017858   0.001101  16.225  < 2e-16 ***
## ratio1      0.002891   0.001740   1.661  0.09662 .  
## ratio2      0.004775   0.001740   2.744  0.00606 ** 
## ratio3      0.004875   0.001740   2.802  0.00509 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.1422 on 50079 degrees of freedom
##   (47 observations deleted due to missingness)
## Multiple R-squared:  0.0002195,  Adjusted R-squared:  0.0001596 
## F-statistic: 3.665 on 3 and 50079 DF,  p-value: 0.01176
```

The intercept is 0.017858, which represents the expected response rate when all the dummy variables are equal to 0/ when there isn't matching.

The coefficient of ratio1 is 0.002891 so when ratio is 1:1, the expected response rate is $0.017858 + 0.002891 = 0.020749$. At a significance level of 0.05, the p-value is 0.09662, greater than 0.05, which means ratio1 is not statistically significant.

The coefficient of ratio2 is 0.004775 so when ratio is 1:2, the expected response rate is $0.017858 + 0.004775 = 0.022633$. At a significance level of 0.05, the p-value is 0.00606, less than 0.05, which means ratio2 is statistically significant.

The coefficient of ratio3 is 0.004875 so when ratio is 1:3, the expected response rate is $0.017858 + 0.004875 = 0.022733$. At a significance level of 0.05, the p-value is 0.00509, less than 0.05, which means ratio3 is statistically significant.

2.3

Calculate the response rate difference between the 1:1 and 2:1 match ratios.

$0.022633 - 0.020749 = 0.001884$

2.4

Based on the regressions you just ran and more generally the results in Table 2a, what do you conclude regarding the effectiveness of using matched donations?

Based on the regression, I see that ratio2 and ratio3 are statistically significant at significance level of 0.05. Ratio3 has the highest coefficient which means it can increase the response rate the most compared to the other two. In table 2a, ratio2 and ratio3 have the same mean. My general conclusion would be, ratio2 and ratio3 can both increase response rate while being statistically significant; so they are both effective.

3. Response rates in red/blue states

3.1

Repeat the regression of response rate on treatment and an intercept (do not include separate match ratio dummies). But this time, base the regression only on respondents in blue states or red states. I.e. run two regressions, one on each of the two sub-samples of data (when defining `data` inside the `lm` command, you can use the following syntax: `data=subset(frame_name, red_state_dummy==0)` and the equivalent expression when red state dummy is equal to one). Interpret the coefficients in both regressions. Is the treatment more effective in red or blue states?

```
reg_3_1_blue = lm(donation_dummy~ treatment, data = subset(df, red_state_dumm
y == 0))
reg_3_1_red = lm(donation_dummy~ treatment, data = subset(df, red_state_dummy
```

```
 == 1))
summary(reg_3_1_blue)
```

```
##
## Call:
## lm(formula = donation_dummy ~ treatment, data = subset(df, red_state_dummy
 ==
##     0))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.02109 -0.02109 -0.02109 -0.02004  0.97996
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.020042   0.001423  14.085   <2e-16 ***
## treatment   0.001043   0.001747   0.597     0.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1425 on 29804 degrees of freedom
## Multiple R-squared:  1.197e-05,  Adjusted R-squared:  -2.159e-05
## F-statistic: 0.3567 on 1 and 29804 DF,  p-value: 0.5504
```

```
summary(reg_3_1_red)
```

```
##
## Call:
## lm(formula = donation_dummy ~ treatment, data = subset(df, red_state_dummy
 ==
##     1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.02339 -0.02339 -0.02339 -0.01459  0.98541
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.014591   0.001737   8.398  < 2e-16 ***
## treatment   0.008802   0.002120   4.152 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1417 on 20240 degrees of freedom
## Multiple R-squared:  0.0008509,  Adjusted R-squared:  0.0008015
## F-statistic: 17.24 on 1 and 20240 DF,  p-value: 3.313e-05
```

The coefficient of treatment of red state is 0.008802, greater than the coefficient of treatment of blue state 0.001043. Treatment is also statistically significant at significance level of 0.05 in red states. In conclusion, treatment is more effective in red states.

3.2

States are of course not randomly assigned. Does the treatment coefficient have a causal interpretation in each of the two regressions? Does the difference in the treatment effect between states have a causal interpretation?

Since the treatment groups are randomly assigned, the treatment coefficient have a causal interpretation in both of the two regressions. However, since the states are not randomly assigned, the difference in the treatment effect between states don't have a causal interpretation.

4. Response rates and donation amount

4.1

Run a regression of dollars given on a treatment dummy and an intercept. Interpret the regression coefficients. Does the treatment coefficient have a causal interpretation?

```
reg_4_1 = lm(donation_amount~ treatment, data =df)
summary(reg_4_1)

##
## Call:
## lm(formula = donation_amount ~ treatment, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -0.97  -0.97  -0.97  -0.81 399.03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.81327    0.06742  12.063   <2e-16 ***
## treatment    0.15361    0.08256   1.861   0.0628 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.709 on 50081 degrees of freedom
##   (47 observations deleted due to missingness)
## Multiple R-squared:  6.911e-05,  Adjusted R-squared:  4.915e-05
## F-statistic: 3.461 on 1 and 50081 DF,  p-value: 0.06282
```

The mean of donation amount of the control group is 0.81327, while that for the treatment group is $0.81327 + 0.15361 = 0.96688$. The treatment coefficient has causal interpretation because it was randomized and is compared with a control group that did not receive a matching offer.

4.2

Next, regress dollars given on a treatment dummy and an intercept, but base the regression only on respondents that made a donation (i.e. donation_dummy is equal to 1). This regression allows you to analyze how much respondents donate *conditional* on donating some positive amount. Interpret the regression coefficients. Does the treatment coefficient have a causal interpretation?

```
reg_4_2 = lm(donation_amount~treatment, data = subset(df, donation_dummy ==1)
)
summary(reg_4_2)
```

```
##
## Call:
## lm(formula = donation_amount ~ treatment, data = subset(df, donation_dummy
 ==
##     1))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43.54 -23.87 -18.87   6.13 356.13
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   45.540      2.423  18.792   <2e-16 ***
## treatment     -1.668      2.872  -0.581    0.561
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.83 on 1032 degrees of freedom
## Multiple R-squared:  0.0003268,  Adjusted R-squared:  -0.0006419
## F-statistic: 0.3374 on 1 and 1032 DF,  p-value: 0.5615
```

The intercept of the regression is 45.540， meaning the average of donation is 45.540 for those who have donated in the control group; while that for the treatment group is $45.540 - 1.668 = 43.872$.

The regression is done on a condition that people have donated; therefore, random assignment doesn't exist here. Consequently, there isn't a casual interpretation.