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.

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

```
##
## Call:
## lm(formula = months_since_last_donation ~ treatment, data = data)
##
## 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
```

ANSWER: The intercept in this regression tells us that those in the control group (i.e. those with the `treatment` dummy equal to zero) on average have gone roughly 13 (12.99814) months since last donating. The coefficient on the `treatment` dummy in the above regression is 0.01369, indicating that those exposed to the treatment on average have gone 0.01369 months longer since last donating. This is consistent with the group-specific means in Table 1 of the paper, which report that the treatment and control groups have gone an average of 13.012 and 12.998 months since last donating, respectively. Taking the difference of these numbers yields 0.014, which is roughly equal to the aforementioned coefficient on the `treatment` dummy, as expected.

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?

ANSWER: The p-value of 0.905 on the `treatment` dummy coefficient implies that this relationship is not statistically significant. This is what we would expect since individuals were assigned to treatment and control groups at random in the experiment.

1.3

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

ANSWER: One take-away from Table 1 is that many of the characteristics measured are similar among each of the groups (all, treatment, and control). Additionally, the standard deviations of each of the group-specific means is very large relative to these means. This demonstrates that there is little relationship between each of these characteristics and whether members were assigned to a treatment or control group (i.e. random assignment was upheld).

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.

```r
summary(lm(donation_dummy ~ treatment, data = data))
```

```
##
## Call:
## lm(formula = donation_dummy ~ treatment, data = data)
##
## 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
```

ANSWER: Here, the response and predictor variables are both dummy variables, making this a linear probability model. The intercept of 0.017858 thus is a measure of the probability of donation for those in the control group (i.e. those with the `treatment` dummy equal to zero). The coefficient on the `treatment` dummy represents that this probability of donation is on average 0.004180 higher for those in the treatment group. These results are consistent with the results in first row of Table 2a.

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.

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

```
##
## Call:
## lm(formula = donation_dummy ~ ratio1 + ratio2 + ratio3, data = data)
##
## 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
```

ANSWER: Again, this is a linear probability model (though it is a multivariate model in this case) and thus the intercept of 0.017858 is a measure of the probability of donation for those in the control group.

The coefficient of 0.002891 on `ratio1` represents the increase in response rate demonstrated by those in the treatment group with a 1:1 match ratio.

The coefficient of 0.004775 on `ratio2` represents the increase in response rate demonstrated by those in the treatment group with a 2:1 match ratio.

The coefficient of 0.004875 on `ratio3` represents the increase in response rate demonstrated by those in the treatment group with a 3:1 match ratio.

2.3

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

ANSWER: To calculate this, we subtract the corresponding coefficients from the regression in 2.2 from one another: 0.004775 - 0.002891 = 0.001884.

The difference in the response rate between the 1:1 and 2:1 match ratios is thus 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?

ANSWER: From the above regressions and the results in Table 2a, we can conclude that matched donations generally tend to increase the response rate to donation requests, but no single price ratio stands out over the others. Increasing the matching ratio from `ratio1` to `ratio2` increases the conversion rate substantially; however, there is little difference between `ratio2` and `ratio3`.

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. Interpret the coefficients in both regressions. Is the treatment more effective in red or blue states?

```r
data_blue = subset(data, red_state_dummy == 0)
outlm_blue = lm(donation_dummy ~ treatment, data = data_blue)
summary(outlm_blue)
```

```
##
## Call:
## lm(formula = donation_dummy ~ treatment, data = data_blue)
##
## 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
```

Here, the intercept of 0.020042 is a measure of the response rate for those in the control group in blue states. The coefficient on the `treatment` dummy shows that this response rate is on average 0.001043 higher for those in the treatment group.

```r
data_red = subset(data, red_state_dummy == 1)
outlm_red = lm(donation_dummy ~ treatment, data = data_red)
summary(outlm_red)
```

```
##
## Call:
## lm(formula = donation_dummy ~ treatment, data = data_red)
##
## 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
```

Here, the intercept of 0.014591 is a measure of the response rate for those in the control group in red states. The coefficient on the `treatment` dummy shows that this response rate is on average 0.008802 higher for those in the treatment group.

ANSWER: From the above regressions, we can see that those in blue states have a higher response rate in the control group, as demonstrated by the intercepts. The effectiveness of the treatment, however, is higher in red states, as demonstrated by the coefficients on the respective treatment dummy variables.

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?

ANSWER: Yes, sub-sample treatment coefficients do have a causal interpretation because treatment is still random in each sub-sample. However, the difference in treatment effects cannot be causally attributed to the political leaning of states because political orientation is not randomly assigned.

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?

```r
summary(lm(donation_amount ~ treatment, data = data))
```

```
##
## Call:
## lm(formula = donation_amount ~ treatment, data = data)
##
## 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
```

ANSWER: The intercept of 0.81327 means that those in the control group on average donated \$0.81327, while the treatment coefficient of 0.15361 implies that those in the treatment group on average donated \$0.15361 more.

The treatment coefficient does have a causal interpretation since the regression is based on the full sample where treatment was randomly assigned.

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?

```r
data_donated = subset(data, donation_dummy == 1)
outlm_donated = lm(donation_amount ~ treatment, data  = data_donated)
summary(outlm_donated)
```

```
##
## Call:
## lm(formula = donation_amount ~ treatment, data = data_donated)
##
## 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
```

ANSWER: The intercept of 45.540 means that those in the control group that did donate some positive amount in the first place on average donated $45.540, while the treatment coefficient of -1.668 implies that those in the treatment group that did donate some positive amount in the first place on average donated $1.668 less than those in the control group.

Treatment does not have a causal interpretation because the respondents select whether to donate or not. Therefore, treatment is not randomly assigned within this sub-sample.