PRESCRIPTIVE MODELS AND DATA ANALYTICS

Problem Set #1

Arnav Garg (906310841)

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).

1 Table 1

Question 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.

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import sys
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

In [ ]:
```python
# Load dataset
charitable_giving = pd.read_csv('charitable_giving.csv')

# Print the number of rows and columns
print(charitable_giving.shape)

# Print the first few rows
charitable_giving.head()
```

```
(50130, 12)
```

Out[ ]:

	donation_amount	donation_dummy	control	treatment	match_ratio	ratio1	ratio2	ratio3	red_state_dummy	months_since_last_dona
0	0.0	0.0	0.0	1.0	1.0	1	0.0	0.0	1.0	
1	0.0	0.0	1.0	0.0	0.0	0	0.0	0.0	1.0	
2	0.0	0.0	1.0	0.0	0.0	0	0.0	0.0	1.0	
3	0.0	0.0	0.0	1.0	3.0	0	0.0	1.0	0.0	
4	0.0	0.0	0.0	1.0	2.0	0	1.0	0.0	0.0	

In [ ]:
```python
model = ols('months_since_last_donation ~ treatment', data=charitable_giving).fit()
print(model.summary())
```

```
                            OLS Regression Results
======================================================================================
Dep. Variable:     months_since_last_donation   R-squared:                       0.000
Model:                                    OLS   Adj. R-squared:                 -0.000
Method:                         Least Squares   F-statistic:                   0.01428
Date:                        Sat, 27 Jan 2024   Prob (F-statistic):              0.905
Time:                                22:26:55   Log-Likelihood:            -1.9585e+05
No. Observations:                       50082   AIC:                         3.917e+05
Df Residuals:                           50080   BIC:                         3.917e+05
Df Model:                                   1
Covariance Type:                    nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     12.9981      0.094    138.979      0.000      12.815      13.181
treatment      0.0137      0.115      0.119      0.905      -0.211       0.238
==============================================================================
Omnibus:                     8031.352   Durbin-Watson:                   1.714
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            12471.135
Skew:                           1.163   Prob(JB):                         0.00
Kurtosis:                       3.751   Cond. No.                         3.23
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The control group's months since last donation were 12.998 months and the treatment group's months since last donation were 13.012 months. The difference is not a lot. These values are similar to what we see in the group-specific means from table 1 of the paper.

Question 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?

The difference between treatment and control is not statistically significant because the p-value is greater than 0.05. This result was expected because treatment was randomly assigned.

Question 3. More generally, describe the take-away from Table 1 in the paper.

Table 1 in the paper tells us that the treatment and control groups were randomly assigned and the difference between the values of variables between these two groups is not statistically significant.

2 Response rate regressions

Question 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.

In [ ]:
```
#Run a linear regression of response rate (the donation dummy) on the treatment dummy (and an intercept)
model = ols('donation_dummy ~ treatment', data=charitable_giving).fit()
print(model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:         donation_dummy   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     9.618
Date:                Sat, 27 Jan 2024   Prob (F-statistic):            0.00193
Time:                        22:26:56   Log-Likelihood:                 26630.
No. Observations:               50083   AIC:                        -5.326e+04
Df Residuals:                   50081   BIC:                        -5.324e+04
Df Model:                           1                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0179      0.001     16.225      0.000       0.016       0.020
treatment      0.0042      0.001      3.101      0.002       0.002       0.007
==============================================================================
Omnibus:                    59814.280   Durbin-Watson:                   1.997
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4317152.727
Skew:                           6.740   Prob(JB):                         0.00
Kurtosis:                      46.440   Cond. No.                         3.23
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The mean response rate for control group is 0.018 and for treatment group is 0.022. These values are similar to what we see in the first row of table 2a of the paper.

Question 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.

In [ ]:
```
#Run a regression on three dummies for match ratio treatment (1:1, 2:1, and 3:1 and an intercept).
model = ols('donation_dummy ~ ratio1 + ratio2 + ratio3', data = charitable_giving).fit()
print(model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:          donation_dummy   R-squared:                       0.000
Model:                             OLS   Adj. R-squared:                  0.000
Method:                  Least Squares   F-statistic:                     3.665
Date:                 Sat, 27 Jan 2024   Prob (F-statistic):             0.0118
Time:                         22:26:56   Log-Likelihood:                 26630.
No. Observations:                50083   AIC:                        -5.325e+04
Df Residuals:                    50079   BIC:                        -5.322e+04
Df Model:                            3                                         
Covariance Type:             nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0179      0.001     16.225      0.000       0.016       0.020
ratio1         0.0029      0.002      1.661      0.097      -0.001       0.006
ratio2         0.0048      0.002      2.744      0.006       0.001       0.008
ratio3         0.0049      0.002      2.802      0.005       0.001       0.008
==============================================================================
Omnibus:                    59812.754   Durbin-Watson:                   1.997
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4316693.217
Skew:                           6.740   Prob(JB):                         0.00
Kurtosis:                      46.438   Cond. No.                         4.26
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The intercept represents the mean response rate for the control group and the coefficients for each of the match ratios represent the difference in response rate for that particular match ratio from the control group.

Question 3. Calculate the response rate difference between the 1:1 and 2:1 match ratios.

In [ ]:
```
# Calculate the response rate difference between the 1:1 and 2:1 match ratios
print("1:1 Match Ratio:", model.params[1])
print("2:1 Match Ratio:", model.params[2])
print("Difference:", model.params[2] - model.params[1])
```

```
1:1 Match Ratio: 0.0028909112451113196
2:1 Match Ratio: 0.004775162266831068
Difference: 0.0018842510217197484
```

Question 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?

Matching is an effective tool to increase number of donations, but it doesn't really affect the number of donations when the match ratio is increased beyond 1:1.

3 Response rates in red/blue states

Question 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?

In [ ]:
```
model = ols('donation_dummy ~ treatment', data=charitable_giving.loc[charitable_giving.red_state_dummy == 1]).fit
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          donation_dummy   R-squared:                       0.001
Model:                             OLS   Adj. R-squared:                  0.001
Method:                  Least Squares   F-statistic:                     17.24
Date:                 Sat, 27 Jan 2024   Prob (F-statistic):           3.31e-05
Time:                         22:26:56   Log-Likelihood:                 10839.
No. Observations:                20242   AIC:                        -2.167e+04
Df Residuals:                    20240   BIC:                        -2.166e+04
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0146      0.002      8.398      0.000       0.011       0.018
treatment      0.0088      0.002      4.152      0.000       0.005       0.013
==============================================================================
Omnibus:                    24251.343   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1766349.071
Skew:                           6.759   Prob(JB):                         0.00
Kurtosis:                      46.721   Cond. No.                         3.25
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [ ]:

```
model = ols('donation_dummy ~ treatment', data=charitable_giving.loc[charitable_giving.red_state_dummy == 0]).fit
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         donation_dummy   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.3567
Date:                Sat, 27 Jan 2024   Prob (F-statistic):              0.550
Time:                        22:26:56   Log-Likelihood:                 15783.
No. Observations:               29806   AIC:                        -3.156e+04
Df Residuals:                   29804   BIC:                        -3.155e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0200      0.001     14.085      0.000       0.017       0.023
treatment      0.0010      0.002      0.597      0.550      -0.002       0.004
==============================================================================
Omnibus:                    35568.600   Durbin-Watson:                   1.996
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          2547856.644
Skew:                           6.727   Prob(JB):                         0.00
Kurtosis:                      46.250   Cond. No.                         3.21
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Treatment is more effective in red states than in blue states because the coefficient for treatment is larger in magnitude in red states than in blue states.

Question 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?

Treatment coefficient has a causal interpretation in each of the two regressions because treatment and control are randomly assigned in each regression. However, difference in treatment effect between states doesn't have a causal interpretation because states aren't randomly assigned.

4 Response rates and donation amount

Question 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?

In [ ]:
```
model = ols('donation_amount ~ treatment', data = charitable_giving).fit()
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        donation_amount   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     3.461
Date:                Sat, 27 Jan 2024   Prob (F-statistic):             0.0628
Time:                        22:26:56   Log-Likelihood:            -1.7946e+05
No. Observations:               50083   AIC:                         3.589e+05
Df Residuals:                   50081   BIC:                         3.589e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.8133      0.067     12.063      0.000       0.681       0.945
treatment      0.1536      0.083      1.861      0.063      -0.008       0.315
==============================================================================
Omnibus:                    96861.113   Durbin-Watson:                   1.987
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        240735713.630
Skew:                          15.297   Prob(JB):                         0.00
Kurtosis:                     341.269   Cond. No.                         3.23
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Control group donates $0.813 on average and treatment group donates $0.967 on average. The treatment effect does have a causal interpretation because treatment and control is assigned randomly on the whole data set.

Question 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?

In [ ]:
```
model = ols('donation_amount ~ treatment', data=charitable_giving.loc[charitable_giving.donation_dummy == 1]).fit
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        donation_amount   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.3374
Date:                Sat, 27 Jan 2024   Prob (F-statistic):              0.561
Time:                        22:26:56   Log-Likelihood:                -5326.8
No. Observations:                1034   AIC:                         1.066e+04
Df Residuals:                    1032   BIC:                         1.067e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     45.5403      2.423     18.792      0.000      40.785      50.296
treatment     -1.6684      2.872     -0.581      0.561      -7.305       3.968
==============================================================================
Omnibus:                      587.258   Durbin-Watson:                   1.838
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5623.279
Skew:                           2.464   Prob(JB):                         0.00
Kurtosis:                      13.307   Cond. No.                         3.49
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Control group donates $45.54 on average and treatment group donates $43.872 on average. The treatment effect does not have a causal interpretation because treatment and control is not assigned randomly on this sample since people make a choice whether to donate or not.