HW4

Nikhil Gopal

10/31/2020

Question 1

```
data = read.csv("progresa.csv")

treatment_average_turnout = mean(data$t2000[data$treatment==1])
control_average_turnout = mean(data$t2000[data$treatment==0])

treatement_average_support = mean(data$pri2000s[data$treatment==1])
control_average_support = mean(data$pri2000s[data$treatment==0])


treatement_average_pri_vote_share = mean(data$pri2000v[data$treatment==1])
control_average_pri_vote_share = mean(data$pri2000v[data$treatment==0])


regression_turnout = lm(data$t2000~data$treatment)
regression_votes_as_share_of_voting_pop = lm(data$pri2000s~data$treatment)

treatment_average_turnout
```

```
## [1] 68.08451
```

```
control_average_turnout
```

```
## [1] 63.81483
```

```
treatement_average_support
```

```
## [1] 38.11145
```

```
control_average_support
```

```
## [1] 34.48895
```

```
summary(regression_turnout)
```

```
##
## Call:
```

```
## lm(formula = data$t2000 ~ data$treatment)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -66.300 -11.700  -1.841   7.778 303.344
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      63.815      2.631  24.255   <2e-16 ***
## data$treatment    4.270      3.216   1.327    0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.91 on 415 degrees of freedom
## Multiple R-squared:  0.004228,   Adjusted R-squared:  0.001829
## F-statistic: 1.762 on 1 and 415 DF,  p-value: 0.1851
```

```
summary(regression_votes_as_share_of_voting_pop)
```

```
##
## Call:
## lm(formula = data$pri2000s ~ data$treatment)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.370 -11.332  -1.673   9.200 104.111
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      34.489      1.554  22.196   <2e-16 ***
## data$treatment    3.622      1.900   1.907   0.0572 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.25 on 415 degrees of freedom
## Multiple R-squared:  0.008686,   Adjusted R-squared:  0.006298
## F-statistic: 3.636 on 1 and 415 DF,  p-value: 0.05722
```

Our average treatment group turnout was about 68%, and control group turnout was about 63%. Our treatment average PRI support as a share of voting population was about 38.1% vs 34.5% for control. The linear regression models really do not make any sense to use and are of little statistical value. They both have near zero R^2 values. This is presumably because the treatment column is really categorical data. The data could have been coded as strings of "treatment" vs and "control" as opposed to 0/1. However, our regression models actually predict quite accurately the turnout and PRI party support, all of the estimates are within 1 of their actual mean values. These results do support the hypothesis that the program mobilized support and turnout, since both average values were higher in precints that were in the treatment group compared to the control group.

Question 2

```
regression_turnout_using_original_model = lm(t2000~treatment +avgpoverty+pobtot1994+votos1994+pri1994+pa
regression_support_using_original_model = lm(pri2000s~treatment +avgpoverty+pobtot1994+votos1994+pri1994

summary(regression_turnout_using_original_model)
```

```
##
## Call:
## lm(formula = t2000 ~ treatment + avgpoverty + pobtot1994 + votos1994 +
##     pri1994 + pan1994 + prd1994, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.173 -11.924  -2.938   6.972 302.183
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 64.0117350 17.3115236   3.698 0.000247 ***
## treatment    4.5494445  3.1346655   1.451 0.147454
## avgpoverty   0.3102553  3.5223779   0.088 0.929855
## pobtot1994  -0.0012128  0.0002316  -5.236 2.63e-07 ***
## votos1994   -0.0261518  0.0354456  -0.738 0.461059
## pri1994      0.0360555  0.0409173   0.881 0.378739
## pan1994      0.0265376  0.0595018   0.446 0.655836
## prd1994      0.0175753  0.0426669   0.412 0.680615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.95 on 409 degrees of freedom
## Multiple R-squared:  0.0785, Adjusted R-squared:  0.06273
## F-statistic: 4.978 on 7 and 409 DF,  p-value: 2.01e-05
```

```
summary(regression_support_using_original_model)
```

```
##
## Call:
## lm(formula = pri2000s ~ treatment + avgpoverty + pobtot1994 +
##     votos1994 + pri1994 + pan1994 + prd1994, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.297  -9.854  -1.322   6.468  94.918
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.9500862  9.4239516   4.027 6.73e-05 ***
## treatment    2.9277395  1.7064319   1.716  0.08697 .
## avgpoverty   0.5329801  1.9174926   0.278  0.78119
## pobtot1994  -0.0004996  0.0001261  -3.962 8.77e-05 ***
## votos1994   -0.0417278  0.0192957  -2.163  0.03116 *
## pri1994      0.0624589  0.0222744   2.804  0.00529 **
## pan1994     -0.0487349  0.0323913  -1.505  0.13321
## prd1994     -0.0287363  0.0232268  -1.237  0.21672
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.3 on 409 degrees of freedom
## Multiple R-squared:  0.2206, Adjusted R-squared:  0.2073
## F-statistic: 16.54 on 7 and 409 DF,  p-value: < 2.2e-16
```

According to our model for turnout, all of the other factors have almost 0 effect on turnout, except for treatment which has a coefficient of about 4.55.

For the model for share of voting population, treatment had a lower coefficient of about 2.92. Average poverty had a noticeably higher effect in this model of about 0.53, while the other factors were mostly negligible.

These results are only slightly different than what was obtained in the previous question, but the most important factor in the models remains the treatment variable.

Question 3

```
#Question 3
natural_turnout_model = lm(t2000~treatment+log(pobtot1994)+t1994+pri1994s+pan1994s+prd1994s+avgpoverty,
natural_support_model = lm(pri2000s~treatment+log(pobtot1994)+t1994+pri1994s+pan1994s+prd1994s+avgpover

summary(natural_turnout_model)
```

```
##
## Call:
## lm(formula = t2000 ~ treatment + log(pobtot1994) + t1994 + pri1994s +
##     pan1994s + prd1994s + avgpoverty, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -119.044   -7.248   -1.180    5.579  176.967
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      19.7984    14.4292   1.372  0.17078
## treatment        -0.1530     1.8182  -0.084  0.93299
## log(pobtot1994)  -3.2471     1.1678  -2.780  0.00568 **
## t1994             0.6605     0.1292   5.112 4.91e-07 ***
## pri1994s          0.1943     0.1371   1.417  0.15720
## pan1994s          0.6374     0.2121   3.005  0.00282 **
## prd1994s          0.3065     0.1471   2.084  0.03780 *
## avgpoverty        2.8621     1.8954   1.510  0.13180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.31 on 409 degrees of freedom
## Multiple R-squared:  0.6921, Adjusted R-squared:  0.6868
## F-statistic: 131.3 on 7 and 409 DF,  p-value: < 2.2e-16
```

```
summary(natural_support_model)
```

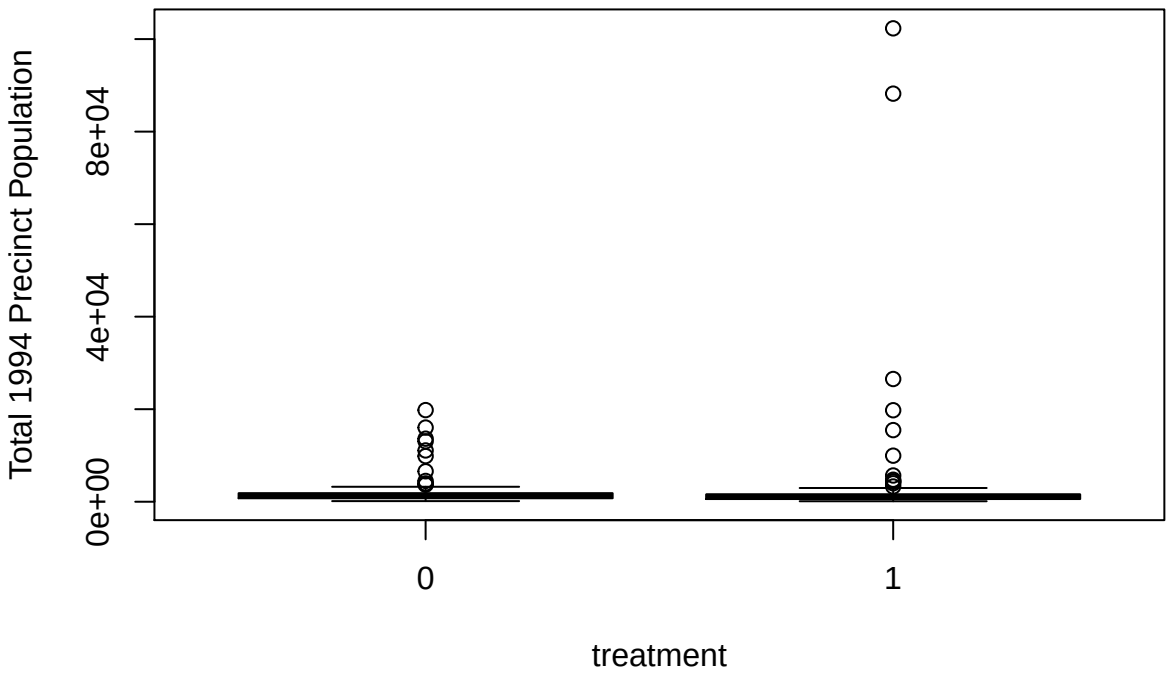
```
##
## Call:
## lm(formula = pri2000s ~ treatment + log(pobtot1994) + t1994 +
##     pri1994s + pan1994s + prd1994s + avgpoverty, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -61.330  -7.237  -0.430   6.116  55.869
##
```

```
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     35.85174    9.98288   3.591 0.000369 ***
## treatment        0.23547    1.25790   0.187 0.851602
## log(pobtot1994) -4.62934    0.80798  -5.730 1.96e-08 ***
## t1994            0.03257    0.08939   0.364 0.715792
## pri1994s         0.51047    0.09488   5.380 1.26e-07 ***
## pan1994s        -0.18384    0.14676  -1.253 0.211052
## prd1994s        -0.05293    0.10175  -0.520 0.603192
## avgpoverty       2.47163    1.31133   1.885 0.060161 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.98 on 409 degrees of freedom
## Multiple R-squared:  0.5794, Adjusted R-squared:  0.5722
## F-statistic: 80.48 on 7 and 409 DF,  p-value: < 2.2e-16
```

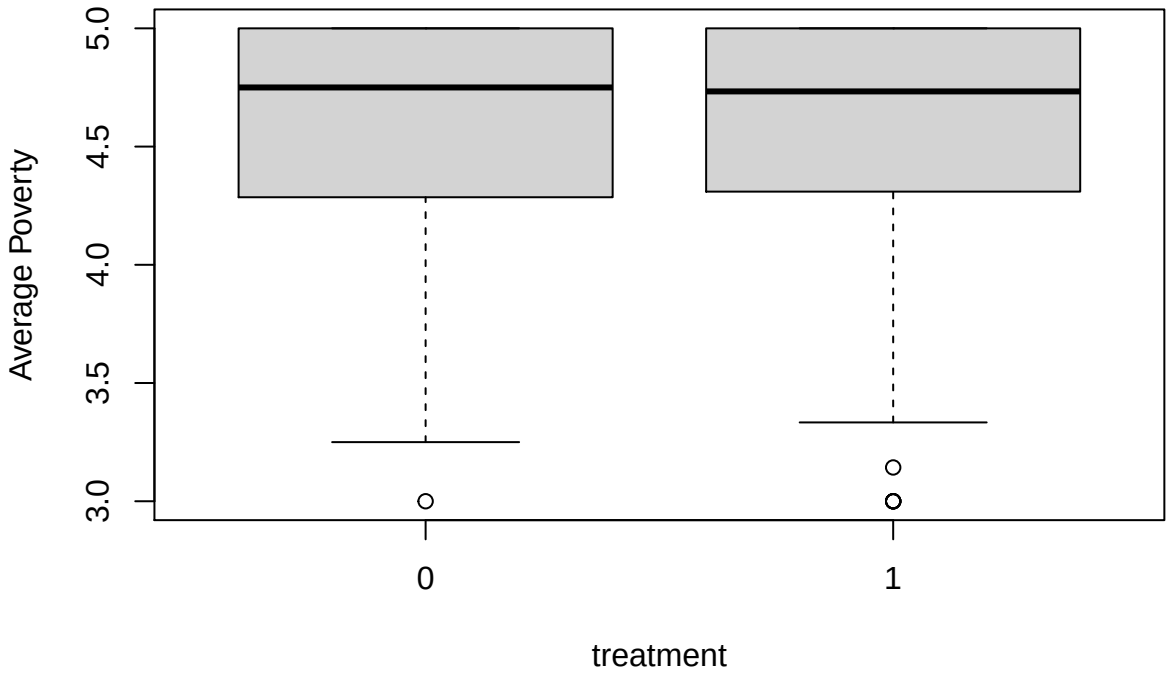
These results are different than what was obtained in the previous model. Population and average poverty have a significant effect in these new models, and treatment has a much smaller influence in these ones (coefficient of -0.15 for turnout and 0.23 for vote share). These natural models fit the data much better, as they had R^2 values of 0.69 and 0.57 compared to 0.06 and 0.2.

Question 4

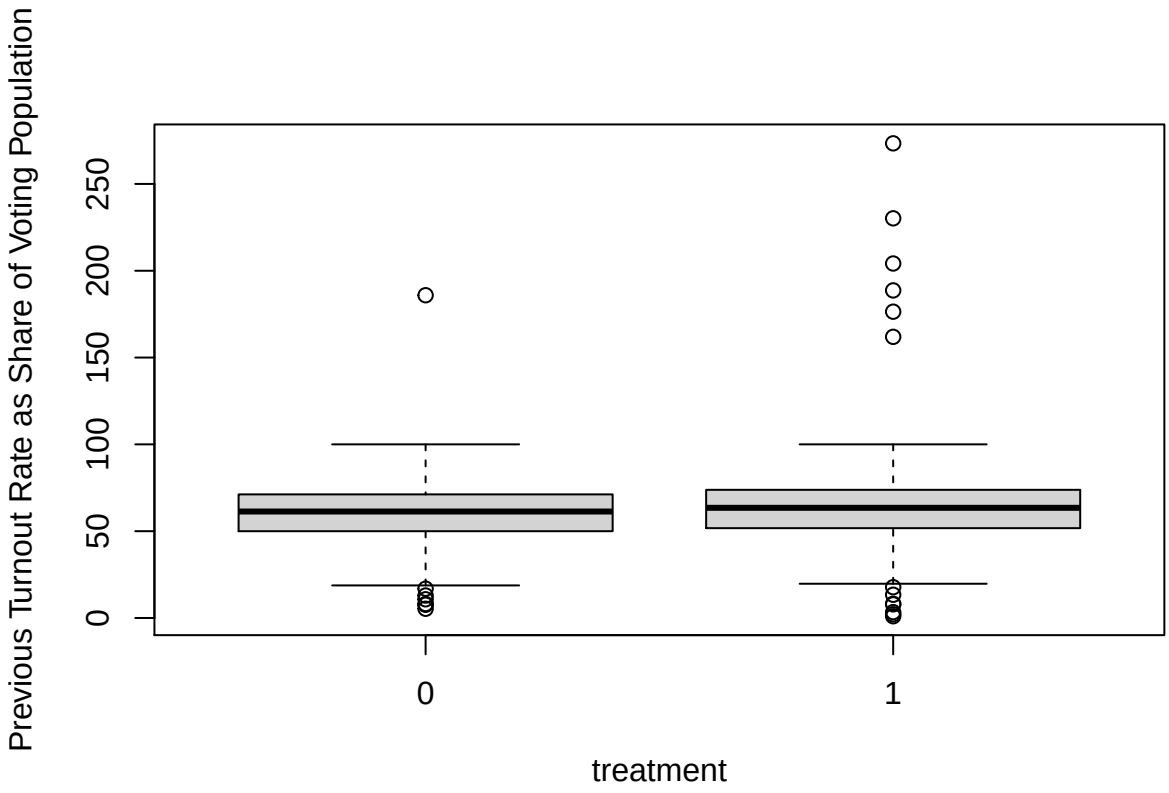
```r
#Question 4

#Total population
boxplot(pobtot1994~treatment, data = data, ylab = "Total 1994 Precinct Population")
```



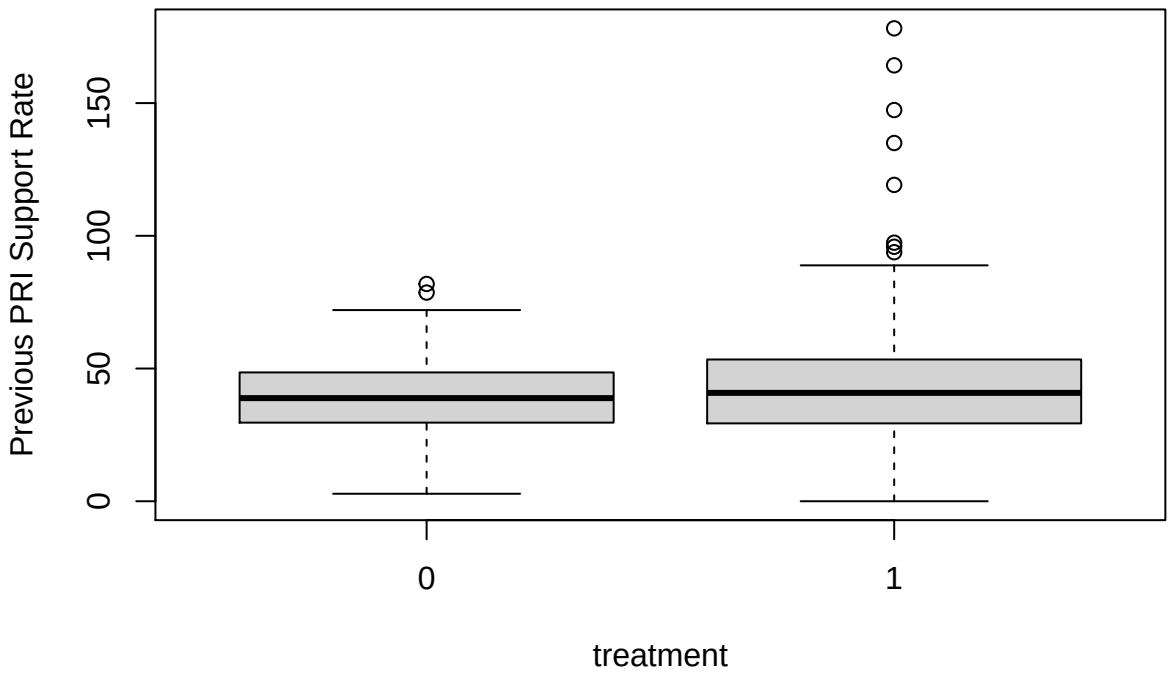

```
#Avg Poverty
boxplot(avgpoverty~treatment, data = data, ylab = "Average Poverty")
```




```
#Previous Turnout Rate
boxplot(t1994~treatment, data = data, ylab = "Previous Turnout Rate as Share of Voting Population")
```




```
#Previous PRI support rate
boxplot(pri1994s~treatment, data=data, ylab = "Previous PRI Support Rate")
```



It appears that many of the distributions are similar. The treatment group had a wider range of support rates, and had slightly more outliers in most of the distribution. Other than that, medians and IQRs were similar across distributions. This is good because it means that the two groups were likely similar and thus we can use this data to make infrences.

Question 5

```
#Question 5

turnout_rate_as_outcome = lm(t2000r ~ treatment + avgpoverty + log(pobtot1994) + t1994r + pri1994v + pa
summary(turnout_rate_as_outcome)
```

```
##
## Call:
## lm(formula = t2000r ~ treatment + avgpoverty + log(pobtot1994) +
##     t1994r + pri1994v + pan1994v + prd1994v, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.041  -4.555   0.029   5.010  29.069
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     33.47237    8.85553   3.780 0.000180 ***
## treatment       -1.08094    0.81683  -1.323 0.186465
## avgpoverty      -0.27734    0.88768  -0.312 0.754874
```

```
## log(pobtot1994) -0.27878    0.47487  -0.587 0.557487
## t1994r           0.22238    0.03102   7.168 3.59e-12 ***
## pri1994v         0.12473    0.05489   2.272 0.023586 *
## pan1994v         0.27356    0.07257   3.770 0.000188 ***
## prd1994v         0.11323    0.05790   1.955 0.051209 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.795 on 408 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1809, Adjusted R-squared:  0.1669
## F-statistic: 12.87 on 7 and 408 DF,  p-value: 5.603e-15
```

```
votes_cast_as_outcome = lm(pri2000v ~ treatment + avgpoverty + log(pobtot1994) + t1994r + pri1994v + pa
summary(votes_cast_as_outcome)
```

```
##
## Call:
## lm(formula = pri2000v ~ treatment + avgpoverty + log(pobtot1994) +
##     t1994r + pri1994v + pan1994v + prd1994v, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.876  -7.686   1.072   7.461  34.758
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     50.22721   14.43904   3.479 0.000558 ***
## treatment        0.80172    1.33185   0.602 0.547536
## avgpoverty       3.19082    1.44737   2.205 0.028042 *
## log(pobtot1994) -2.59489    0.77428  -3.351 0.000879 ***
## t1994r          -0.08612    0.05058  -1.703 0.089421 .
## pri1994v         0.35738    0.08950   3.993 7.73e-05 ***
## pan1994v        -0.48863    0.11832  -4.130 4.41e-05 ***
## prd1994v        -0.24158    0.09441  -2.559 0.010862 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.71 on 408 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4836, Adjusted R-squared:  0.4747
## F-statistic: 54.58 on 7 and 408 DF,  p-value: < 2.2e-16
```

When using turnout as the treatment variable, the treatment reduces turnout (evidenced by negative coefficient in the model), suggesting that the cash transfer program isn't effective in increasing turnout. With respect to vote share, treatment still increases PRI vote share, but other variables (average poverty and population) appear to affect vote share more than the treatment, suggesting that the cash transfer program isn't too effective on increasing PRI vote share either.