Empirical Project 1

Ben Noon

Load Library and Data

```r
library(pacman)
p_load(tidyverse, stargazer, data.table, tidyfast, dtplyr, haven, modelsummary,collapse)

ohp = read_dta("ohp.dta")
ohp[] = lapply(ohp, as.numeric)
ohp_dt <- as.data.table(ohp)
```

Questions

Q1

Explain the difference between the variables treatment and ohp_all_ever_survey. Explain why treatment is the treatment variable (Di), rather than ohp_all_ever_survey.

Treatment is when someone either won the OHP lottery or did not win the lottery (1, 0).

ohp_all_ever_survey is if the person was every enrolled into medicaid.

Treatment is the variable we want because we are looking to see if increased access to medicaid improves health outcomes. ohp_all_ever_survey is independent of the research question and gives us a look at the compliance rate.

Q2

Provide evidence that the OHP lottery really did randomly assign individuals to treatment and control groups. Similar to Table 1 in Taubman et al. (2014), please create a nicely formatted table that reports means of 4 to 6 relevant characteristics for individuals in the control group

```r
ohp_cntrl <- subset(ohp_dt, treatment ==0, select =
c(age_inp, race_black_inp, race_white_inp, race_nwother_inp, gender_inp, edu_inp))

summary(ohp_cntrl)
```

```
    age_inp      race_black_inp   race_white_inp   race_nwother_inp
 Min.   :19.00   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:30.00   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :41.00   Median :0.0000   Median :1.0000   Median :0.0000  
 Mean   :40.61   Mean   :0.1073   Mean   :0.6898   Mean   :0.1425  
 3rd Qu.:50.00   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.0000  
 Max.   :68.00   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
 NA's   :1       NA's   :16       NA's   :16       NA's   :16      
   gender_inp        edu_inp     
 Min.   :0.0000   Min.   :1.000  
 1st Qu.:0.0000   1st Qu.:2.000  
 Median :1.0000   Median :2.000  
 Mean   :0.5688   Mean   :2.238  
 3rd Qu.:1.0000   3rd Qu.:3.000  
 Max.   :2.0000   Max.   :4.000  
                  NA's   :3      
```

Q3

For each of the variables you summarized above, calculate:
(i) the difference between the mean in the treatment group and the mean in the control group;2
(ii) the standard error for the difference in means.

```r
age <- lm(data = ohp_dt, age_inp ~ treatment)
edu <- lm(data = ohp_dt, edu_inp ~ treatment)
black <- lm(data = ohp_dt, race_black_inp~treatment)
white <- lm(data = ohp_dt, race_white_inp~treatment)
other <- lm(data = ohp_dt, race_nwother_inp~treatment)
gen <- lm(data = ohp_dt, gender_inp ~ treatment)


msummary(list(edu, age, white, black, other, gen), stars = TRUE, title = "Balance Check")
```

Table 1: Balance Check

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	2.238***	40.606***	0.690***	0.107***	0.142***	0.569***
	(0.012)	(0.153)	(0.006)	(0.004)	(0.005)	(0.006)
treatment	0.022	0.380+	−0.003	−0.008	0.003	−0.006
	(0.016)	(0.212)	(0.008)	(0.006)	(0.006)	(0.009)
Num.Obs.	12 218	12 228	12 190	12 190	12 190	12 229
R2	0.000	0.000	0.000	0.000	0.000	0.000
R2 Adj.	0.000	0.000	0.000	0.000	0.000	0.000
AIC	32 318.9	94 851.5	15 836.9	5594.0	9095.8	17 552.6
BIC	32 341.1	94 873.8	15 859.1	5616.2	9118.0	17 574.8
Log.Lik.	−16 156.445	−47 422.765	−7915.432	−2793.996	−4544.903	−8773.298
F	1.737	3.225	0.128	1.924	0.277	0.463
RMSE	0.91	11.70	0.46	0.30	0.35	0.50

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Instead of adding the coefficients and standard errors to my earlier graph, I used the modelsummary package to created a better looking graph that contains all the information. The intercept is the mean of the control group for each variable.

Q4

Is the balance table consistent with individuals having been randomly assigned to treatment group and control groups? Why or why not?

I would argue that the balance table shows individuals were randomly assigned to treatment and control groups. None of the coefficients are statistically significant which would indicate one control variable was different from the control.

Q5

Estimate the compliance rate for the OHP experiment. That is, what is the effect of being assigned to the treatment group on the probability of being enrolled in Medicaid?

```
mod1 = lm(ohp_all_ever_survey~ treatment + age_inp + edu_inp + race_black_inp + race_white

msummary(mod1, vcov = "HC1")
```

	Model 1
(Intercept)	0.151
	(0.018)
treatment	0.256
	(0.008)
age_inp	−0.001
	(0.000)
edu_inp	−0.025
	(0.004)
race_black_inp	0.085
	(0.015)
race_white_inp	0.038
	(0.010)
gender_inp	0.089
	(0.008)
Num.Obs.	12 187
R2	0.091
R2 Adj.	0.091
AIC	14 197.4
BIC	14 256.6
Log.Lik.	−7090.691
F	221.099
Std.Errors	HC1

The intercept is at 15.8% and the coefficient is at 25.4%. Averaging these two numbers show the over all compliance rate of ~20.6%.

Q6

What is the intent-to-treat (ITT) effect of the OHP experiment on health outcomes? Please create a nicely formatted table that reports ITT estimates on 4 to 6 relevant health outcomes. Again, part of this question is to get you to think about which 4 to 6 variables could be used as health outcome variables.

```
models = list(
"Depression" = lm(dep_dx_post_lottery ~ treatment + age_inp + edu_inp + race_black_inp + r
"Diabetes" = lm(dia_dx_post_lottery ~ treatment + age_inp + edu_inp + race_black_inp + rac
"Hypertension" = lm(hbp_dx_post_lottery ~ treatment + age_inp + edu_inp + race_black_inp +
"Perscription Num" = lm(rx_num_mod_inp ~ treatment + age_inp + edu_inp + race_black_inp +
)


msummary(models, stars = TRUE, title = "List of Models")
```

Q7

What is the "treatment on the treated" effect (ATET) of the OHP experiment, i.e. the effect among those who applied for Medicaid? Estimate it for every health outcome you chose in question 6 and provide some intuition for the calculation of this estimate.

Table 2: List of Models

	Depression	Diabetes	Hypertension	Perscription Num
(Intercept)	0.064***	−0.014**	−0.002	−2.224***
	(0.009)	(0.005)	(0.010)	(0.116)
treatment	0.005	0.008***	0.002	0.111*
	(0.004)	(0.002)	(0.004)	(0.050)
age_inp	0.000*	0.001***	0.002***	0.079***
	(0.000)	(0.000)	(0.000)	(0.002)
edu_inp	−0.008***	−0.002	−0.011***	−0.031
	(0.002)	(0.001)	(0.002)	(0.028)
race_black_inp	0.007	−0.003	0.013+	0.704***
	(0.007)	(0.004)	(0.008)	(0.091)
race_white_inp	0.009+	−0.009**	0.002	0.735***
	(0.005)	(0.003)	(0.005)	(0.061)
gender_inp	0.022***	0.006**	−0.005	0.599***
	(0.004)	(0.002)	(0.004)	(0.050)
Num.Obs.	12 056	12 146	11 907	11 876
R2	0.004	0.008	0.012	0.127
R2 Adj.	0.003	0.007	0.012	0.126
AIC	−2194.6	−15 688.0	−897.0	57 337.0
BIC	−2135.4	−15 628.8	−838.0	57 396.1
Log.Lik.	1105.287	7851.999	456.520	−28 660.498
F	7.821	15.784	24.135	287.190
RMSE	0.22	0.13	0.23	2.70

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```r
atet_dep= lm(dep_dx_post_lottery ~ treatment + age_inp + edu_inp + race_black_inp + race_w

atet_dia = lm(dia_dx_post_lottery ~ treatment + age_inp + edu_inp + race_black_inp + race_

atet_hyp = lm(hbp_dx_post_lottery ~ treatment + age_inp + edu_inp + race_black_inp + race_

atet_rx = lm(rx_num_mod_inp ~ treatment + age_inp + edu_inp + race_black_inp + race_white_


models2 = list(
  "ATET Depression" = atet_dep$coefficients[2]/mod1$coefficients[2],
  "ATET Diabetes" = atet_dia$coefficients[2]/mod1$coefficients[2],
  "ATET Hypertension" = atet_hyp$coefficients[2]/mod1$coefficients[2],
  "ATET Perscription Number" = atet_rx$coefficients[2]/mod1$coefficients[2]
)

models2
```

```
$`ATET Depression`
 treatment
0.01955166

$`ATET Diabetes`
 treatment
0.03216333

$`ATET Hypertension`
  treatment
0.008415786

$`ATET Perscription Number`
treatment
0.4327228
```

The intent to treat coefficients for depression, diabetes, hypertension, and prescription number are 0.005, 0.008, 0.002, and 0.111. I take the compliance rate of 0.25The ATET effects are 0.0195, 0.0322, 0.0084, and 0.4327.

Q8

Do you have to worry about attrition bias in analyzing this data? Explain why or why not.

I am worried about attrition bias. Since attrition bias is where individuals non-randomly break from treatment or control guidelines, my guess would be that people of certain income strata would break more than other people from a different income strata. Someone of middle class is more likely to commit attrition in the treatment group over the control group.

Q9

Suppose that you are submitting these results to a general interest journal such as Science for publication. Write an abstract of 200 or fewer words describing what you have found in your analysis of the OHP data, similar to the abstract in Taubman et al. (2014).

Public health spending is an intensely debated topic within the United States. Data from the Oregon Health Insurance experiment was gathered to see if there is a noticeable increase in health outcomes. There are varying degrees of success in improving health outcomes for those who were given access to Medicaid, but due to some missing information, attrition bias cannot be ruled out.