PE-PSet3

Drazzel Feliu - 12174100

For this assignment, provide a write-up where you answer the questions below, selectively cutting and pasting output where needed. Be concise in your write-up; excess wordiness will be penalized. Also, submit a log file that includes commands and results for your entire analysis. The assignment makes use of AganStarrQJEData.dta, which you can find on Canvas.

In this problem set we will reproduce some of Amanda Agan and Sonja Starr's basic results, so start by reading their paper (Ban the box, criminal records, and racial discrimination: A field experiment), which you can find on Canvas.

```
# load data set
data <- read_dta("AganStarrQJEData.dta")
# create data table identifying class and labels for each variable
datainfo <- data.frame(variable=colnames(data),
          class=sapply(data, class)
          )
label=unlist(lapply(data, function (x) attr(x, "label")))
label <- as.data.frame(label)
label$variable <- rownames(label)
label <- label[c(2,1)]
datainfo <- left_join(datainfo, label, by = "variable")
```

```
## Warning: Column `variable` joining factor and character vector, coercing
## into character vector
```

```
# summary statistics of variables
summary(data)
```

Questions:

1

For this question, restrict your analysis to the set of job applications that asked about criminal records ("Box" applications) in the before period ("pre-BTB" period). (Note: there are some applications that did not have a box in the pre-BTB period, but then added them in the post- period. Agan and Star code these as "remover = -1" in their data and call them "reverse compliers." Exclude these observations from your analysis throughout this assignment.)

```
# Permanently filter data for all reverse compliers
data <- filter(data, remover!=-1)
```

A)

What is the average callback rate for people who committed crimes? For those who didn't? Is the difference statistically significant?

Answer

```
# Summary of Means across both groups
data %>% group_by(crime) %>% summarise(mean = mean(response)*100)
```

```
## # A tibble: 2 x 2
##   crime       mean
##   <dbl+lbl> <dbl>
## 1 0          12.5
## 2 1          10.9
```

```
# Significance test across two groups for the callback rate
t.test(response~crime, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  response by crime
## t = 2.9346, df = 14547, p-value = 0.003345
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.005183336 0.026034902
## sample estimates:
## mean in group 0 mean in group 1
##        0.1250518        0.1094426
```

The average callback rate for people who committed crimes is 10.94% and is 12.51% for individuals who did not commit crimes. The difference is considered statistically significant at a 95% confidence level.

B)

Can we interpret this as a causal effect? Explain briefly.

Answer

Interpretating this relationship as causal currently is premature. We haven't controlled for the type of application individuals are receiving nor across other demographic controls that may have an influence on the rate of callbacks (level of education, interviewer bias through perceptions of race, geographic location of individuals, hiring needs across time). Several variables may disentangle the impact of criminal history on positive responses to applications.

2

Now consider just the "Box" applications but include both the pre- and post-BTB periods.

A)

Regress callback rates on race, GED, and employment gap. Include "chain"1 and "center" fixed effects. Does race appear to have an effect on callback rates? Does this coefficient have a causal interpretation?

```
reg1 <- data %>% lm(response~white + ged + empgap + chain_id + center, data = .)
```

```r
stargazer(reg1, type = "latex", title = "Call Back Rates With Fixed Effects", covariate.labels = c("Rac
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Nov 01, 2018 - 16:17:13

Table 1: Call Back Rates With Fixed Effects

	Dependent variable:
	response
Race = White	0.024113*** (0.005295)
GED Acquired	−0.002239 (0.005296)
Employment Gap Exists	0.000579 (0.005296)
Chain	−0.000070** (0.000030)
Center	0.000539*** (0.000054)
Constant	0.082800*** (0.008073)
Observations	14,637
R^2	0.008712
Adjusted R^2	0.008374
Residual Std. Error	0.320283 (df = 14631)
F Statistic	25.718140*** (df = 5; 14631)
Note:	*p<0.1; **p<0.05; ***p<0.01

B)

Estimate the model again, but without the chain and center fixed effects. Does the coefficient on "white" change? Why is it important to include chain and center fixed effects?

```r
reg2 <- data %>% lm(response~white + ged + empgap, data = .)
```

```r
stargazer(reg2, type = "latex", title = "Call Back Rates Without Fixed Effects", covariate.labels = c("
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Nov 01, 2018 - 16:17:13

Table 2: Call Back Rates Without Fixed Effects

	Dependent variable:
	response
Race = White	0.024121*** (0.005314)
GED Acquired	−0.001141 (0.005314)
Employment Gap Exists	0.001278 (0.005314)
Constant	0.105053*** (0.005299)
Observations	14,637
R^2	0.001413
Adjusted R^2	0.001209
Residual Std. Error	0.321438 (df = 14633)
F Statistic	6.903060*** (df = 3; 14633)
Note:	*p<0.1; **p<0.05; ***p<0.01

C)

Now add the "conviction" variable. What happens to the coefficient on "white"? If the coefficient changes, does this mean that the previous regression was subject to omitted variable bias?

```
reg3 <- data %>% lm(response~white + ged + empgap + crime, data = .)
```

```
stargazer(reg3, type = "latex", title = "Call Back Rates With Crime", covariate.labels = c("Race = White
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Nov 01, 2018 - 16:17:13

Table 3: Call Back Rates With Crime

	Dependent variable:
	response
Race = White	0.024474*** (0.005314)
GED Acquired	−0.001378 (0.005313)
Employment Gap Exists	0.001164 (0.005313)
Crime Committed	−0.016151*** (0.005314)
Constant	0.113209*** (0.005938)
Observations	14,637
R^2	0.002043
Adjusted R^2	0.001770
Residual Std. Error	0.321347 (df = 14632)
F Statistic	7.489315*** (df = 4; 14632)
Note:	*p<0.1; **p<0.05; ***p<0.01