marshad_HW 1

2023-01-15

```
#FINAL
setwd("/Users/misbaharshad/Downloads/")
df <- read.csv("MTO (1).csv")</pre>
sum(df$treated==1)
sum(df$treated==0)
sum(df$treated==1)+sum(df$treated==0)
#Question 1: Fill in the following covariate balance table.
treated_mean <- mean(df$male_respondant [df$treated== 1])</pre>
control_mean <- mean(df$male_respondant[df$treated==0])</pre>
treated_var <- var(df$male_respondant [df$treated== 1])</pre>
control_var <- var(df$male_respondant[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
######################
treated_mean <- mean(df$black_respondant [df$treated== 1])</pre>
control_mean <- mean(df$black_respondant[df$treated==0])</pre>
treated_var <- var(df$black_respondant [df$treated== 1])</pre>
control_var <- var(df$black_respondant[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
```

```
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/</pre>
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t statistic), df=(size control+size treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#########
treated_mean <- mean(df$high_school_grad_start [df$treated== 1])</pre>
control_mean <- mean(df$high_school_grad_start[df$treated==0])</pre>
treated_var <- var(df$high_school_grad_start [df$treated== 1])</pre>
control_var <- var(df$high_school_grad_start[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#########
treated_mean <- mean(df$never_married_start [df$treated== 1])</pre>
control_mean <- mean(df$never_married_start[df$treated==0])</pre>
treated var <- var(df$never married start [df$treated== 1])</pre>
control_var <- var(df$never_married_start[df$treated==0])</pre>
size treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
```

```
print(pvalue)
########
treated_mean <- mean(df$parent_before_18 [df$treated== 1])</pre>
control_mean <- mean(df$parent_before_18[df$treated==0])</pre>
treated_var <- var(df$parent_before_18 [df$treated== 1])</pre>
control_var <- var(df$parent_before_18[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#########
treated_mean <- mean(df$on_welfare_start [df$treated== 1])</pre>
control_mean <- mean(df$on_welfare_start[df$treated==0])</pre>
treated_var <- var(df$on_welfare_start [df$treated== 1])</pre>
control_var <- var(df$on_welfare_start[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
########
treated_mean <- mean(df$victim_of_crime_start [df$treated== 1])</pre>
control_mean <- mean(df$victim_of_crime_start[df$treated==0])</pre>
treated_var <- var(df$victim_of_crime_start [df$treated== 1])</pre>
control_var <- var(df$victim_of_crime_start[df$treated==0])</pre>
```

```
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#########
treated_mean <- mean(df$verydissat_neighborhood_start [df$treated== 1])</pre>
control_mean <- mean(df$verydissat_neighborhood_start[df$treated==0])</pre>
treated_var <- var(df$verydissat_neighborhood_start [df$treated== 1])</pre>
control_var <- var(df$verydissat_neighborhood_start[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#######
treated_mean <- mean(df$want_move_gangs_drugs_start [df$treated== 1])</pre>
control_mean <- mean(df$want_move_gangs_drugs_start[df$treated==0])</pre>
treated_var <- var(df$want_move_gangs_drugs_start [df$treated== 1])</pre>
control_var <- var(df$want_move_gangs_drugs_start[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
```

```
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
########################
/pagebreak
#Question 2: Discuss your findings. Do these baseline characteristics appear balanced?
#Yes, these results appear balanced because the differences between the treated and controlled
#means are small. We can tell because the numbers in the "difference" columns are small. This means tha
#means is less than our significance level of 0.95.
#Question 3: Calculate the treatment effect of receiving a voucher and the statistical significance for
#of the outcomes, filling out the table below. Discuss your findings -- What do you conclude
#about the effectiveness of the voucher program?
#For both happy and working end, if we compare the p-values we see there is no statistically significan
#recieve vouchers. We fail to reject the null hypothesis stating that there is no statistically
#significant difference after recving the voucher in employment and happiness outcomes.
working_treated <- filter(df$working_end ==1 [df$treated==1])</pre>
treated_mean <- select(df$working_end [df$treated== 1])</pre>
control_mean <- mean(df$working_end[df$treated==0])</pre>
treated_var <- var(df$working_end [df$treated== 1])</pre>
control_var <- var(df$working_end[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
###############
treated_mean <- mean(df$happy_end [df$treated== 1])</pre>
control_mean <- mean(df$happy_end[df$treated==0])</pre>
```

treated_var <- var(df\$happy_end [df\$treated== 1])</pre>

```
control_var <- var(df$happy_end[df$treated==0])</pre>
size_treated <- sum(df$treated==1)</pre>
size_control <- sum(df$treated==0)</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#Q4: Calculate the treatment effect and statistical significance of receiving a voucher on working
#at endline, working end, and happiness at endline, happy end, separately for each city, filling
#out the table below. Discuss your results.
#When comparing happy and working end outcomes between cities, there is no statistically significant di
#except in the case of Baltimore-working end. The p-value is less than our significance level of 0.05
#and when the p-value is low, we reject the null. This means that, based on the parameters of expiremen
#alone, the vouchers DID make a difference in employment outcomes in Boston. Other than that, all the o
#cities and outcomes had p-values above 0.05 meaning we fail to reject the null hypothesis stating that
#the voucher has no statistically significant effect.
treated_mean <- mean(df$happy_end [df$city== "Baltimore" & df$treated ==1])</pre>
control_mean <- mean(df$happy_end [df$city== "Baltimore" & df$treated ==0])</pre>
treated_var <- var(df$happy_end [df$city== "Baltimore" & df$treated ==1])</pre>
control_var <- var(df$happy_end [df$city== "Baltimore" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "Baltimore")</pre>
size_control <- sum(df$treated==0 & df$city== "Baltimore")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)
print(pvalue)
##################
treated_mean <- mean(df$working_end [df$city== "Baltimore" & df$treated ==1])
```

```
control_mean <- mean(df$working_end [df$city== "Baltimore" & df$treated ==0])</pre>
treated var <- var(df$working end [df$city== "Baltimore" & df$treated ==1])
control_var <- var(df$working_end [df$city== "Baltimore" & df$treated ==0])</pre>
size treated <- sum(df$treated==1 & df$city== "Baltimore")</pre>
size_control <- sum(df$treated==0 & df$city== "Baltimore")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/</pre>
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#######################
treated_mean <- mean(df$happy_end [df$city== "Boston" & df$treated ==1])</pre>
control_mean <- mean(df$happy_end [df$city== "Boston" & df$treated ==0])</pre>
treated_var <- var(df$happy_end [df$city== "Boston" & df$treated ==1])</pre>
control_var <- var(df$happy_end [df$city== "Boston" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city == "Boston")</pre>
size_control <- sum(df$treated==0& df$city == "Boston")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
treated_mean <- mean(df$working_end [df$city== "Boston" & df$treated ==1])</pre>
control_mean <- mean(df$working_end [df$city== "Boston" & df$treated ==0])</pre>
treated_var <- var(df$working_end [df$city== "Boston" & df$treated ==1])</pre>
control_var <- var(df$working_end [df$city== "Boston" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "Boston")</pre>
size control <- sum(df$treated==0 & df$city== "Boston")</pre>
```

```
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
###############
treated_mean <- mean(df$happy_end [df$city== "Chicago" & df$treated ==1])
control_mean <- mean(df$happy_end [df$city== "Chicago" & df$treated ==0])</pre>
treated_var <- var(df$happy_end [df$city== "Chicago" & df$treated ==1])</pre>
control_var <- var(df$happy_end [df$city== "Chicago" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "Chicago")</pre>
size_control <- sum(df$treated==0 & df$city== "Chicago")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
treated_mean <- mean(df$working_end [df$city== "Chicago" & df$treated ==1])
control_mean <- mean(df$working_end [df$city== "Chicago" & df$treated ==0])</pre>
treated_var <- var(df$working_end [df$city== "Chicago" & df$treated ==1])</pre>
control_var <- var(df$working_end [df$city== "Chicago" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "Chicago")</pre>
size_control <- sum(df$treated==0 & df$city== "Chicago")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
```

```
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#########
treated_mean <- mean(df$happy_end [df$city== "Los Angeles" & df$treated ==1])
control_mean <- mean(df$happy_end [df$city== "Los Angeles" & df$treated ==0])</pre>
treated_var <- var(df$happy_end [df$city== "Los Angeles" & df$treated ==1])</pre>
control_var <- var(df$happy_end [df$city== "Los Angeles" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "Los Angeles")</pre>
size_control <- sum(df$treated==0& df$city== "Los Angeles")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
treated_mean <- mean(df$working_end [df$city== "Los Angeles" & df$treated ==1])
control_mean <- mean(df$working_end [df$city== "Los Angeles" & df$treated ==0])</pre>
treated_var <- var(df$working_end [df$city== "Los Angeles" & df$treated ==1])</pre>
control_var <- var(df$working_end [df$city== "Los Angeles" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "Los Angeles")</pre>
size control <- sum(df$treated==0 & df$city== "Los Angeles")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#################
treated_mean <- mean(df$happy_end [df$city== "New York City" & df$treated ==1])
control_mean <- mean(df$happy_end [df$city== "New York City" & df$treated ==0])</pre>
```

```
treated_var <- var(df$happy_end [df$city== "New York City" & df$treated ==1])
control_var <- var(df$happy_end [df$city== "New York City" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "New York City")</pre>
size control <- sum(df$treated==0 & df$city== "New York City")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)</pre>
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
treated_mean <- mean(df$working_end [df$city== "New York City" & df$treated ==1])
control_mean <- mean(df$working_end [df$city== "New York City" & df$treated ==0])</pre>
treated_var <- var(df$working_end [df$city== "New York City" & df$treated ==1])</pre>
control_var <- var(df$working_end [df$city== "New York City" & df$treated ==0])</pre>
size_treated <- sum(df$treated==1 & df$city== "New York City")</pre>
size_control <- sum(df$treated==0 & df$city== "New York City")</pre>
diff <- treated_mean - control_mean</pre>
denom <- ( (size_treated-1 ) * treated_var + (size_control - 1) * control_var ) * (1/ size_control + 1/
t_statistic <- diff / sqrt(denom)
print(t_statistic)
pvalue <- 2*pt(abs(t_statistic), df=(size_control+size_treated -2), lower.tail = FALSE)</pre>
print(pvalue)
#5. Have a look at the variable moved using voucher end. Did everyone who received a voucher
#move? How does it affect the interpretation of our results if not everyone moves? How does
#the interpretation change if everyone had moved?
#No, a smaller group of people who received the voucher actually moved, meaning the expirement had an #
#intent to treat parameter. If everyone who recieved a voucher moved, we would have perfect compliance.
#Since a small group moved, we can't extrapolate the voucher's effectiveness as accurately.
```

#Q6: As is common in experiments, everyone in the MTO study volunteered to participate. In #what ways might this affect the external validity of the results?

#external validity refers to how well the expirement can be generalized, so if all the participants #are volunteers, we are only getting data from a group who CHOSE to be in the study. The characteritics #group might be very different from the population. For example, if a person CHOSE to participate #in this study, they may already be more likely to move. If the sample characteristics and the #population characteristics are different, it's hard to repeat the expirement elsewhere and expect #similar results.

#Q7: Let's try redoing Part (3) using a simple linear regression.
#For each of the two outcomes in Part (3) above, run a simple linear regression of the outcome
#on a treatment indicator:
working_end_reg <- lm(working_end ~ treated==1, data = df)
summary(working_end_reg)
happy_end_reg <- lm(happy_end ~ treated==1, data=df)
summary(happy_end_reg)</pre>

#the numbers from our regression and from the table in question 3.

#########################

#PART 3: SHORT ANSWER

- #1) INCORRECT, external validity is not more important than internal validity because if the ability to #generalize an expirement is not more important than the integrity of the expirement itslef. If the #expirement design is flawed, it automatically should not be replicated with a different population. Fur #less information can be extrapolated from a non-internally valid expirement but you can still get #helpful information from an expirement that is not able to be replicated under different groups or #ci.
- #2) CORRECT, T-hat is a sample of T (the true population). Once the sample size becomes infinitte, it
- #3) Possible confound: someone who is not vaccinated may also be more likely to engage in other risky b #If this is true, they may be more reckless drivers as well which would make the causality between reck #behaviour and risk of accidents, NOT covid vaccines and accidents.
- #4) Possible confound: People who choose to exercise may also be healthier in other ways, such as eating the substitution and thus having a boosted immune system which is better equipped to recover from covid. In the exercise is not the CAUSE of protecting against fatal covid and the lurking variable may be the health the the substitution of the eating of the health of the substitution of the eating of the health of the substitution of the eating of the health of the substitution of the eating of the

R. Markdown





Winter 2023 Problem Set 1 PPHA 311

1. Fill in the following covariate balance table. In Column (4), calculate the p-value using a two-sample t-test assuming equal variance. The test-statistic for this test is given by: 2

$$t = \frac{\overline{Y}_T - \overline{Y}_C}{\sqrt{\frac{(N_T - 1)s_T^2 + (N_C - 1)s_C^2}{N_T + N_C - 2} \left(\frac{1}{N_C} + \frac{1}{N_T}\right)}}$$

where t is distributed as Student's t with $N_T + N_C - 2$ degrees of freedom. In terms of notation, $\overline{Y_T}$ is the sample mean of the treated group, $\overline{Y_C}$ is the sample mean of the control group, N_T is the sample size of the treated group, N_C is the sample size of the control group, s_T^2 is the sample variance of the treated group, s_C^2 is the sample variance of the control group.

For full credit, please code all parts of your significance test by hand, instead of using a built-in package like t.test in R. (3 points)

	(1)	(2)	(3)	(4)
	Control	Treated	Difference	
Baseline characteristic	Mean	Mean	(2)-(1)	p-value
male_respondant	0.0184	0.0126	-0.0057	0.1898
black_respondant	0.6145	0.6363	0.02178	0.2191
high_school_grad_start	0.3669	0.3809	0.0139	0.4315
never_married_start	0.6147	0.6260	-0.0157	0.5741
parent_before_18	0.2458	0.2572	0.0114	0.4738
$on_welfare_start$	0.7655	0.7553	1010.0-	0.5161
$victim_of_crime_start$	0.4073	0.4254	1810.0	0.3171
verydissat_neighborhood_start	0.4495	0.4723	6.0228	0.2122
want_move_gangs_drugs_start	0.7848	0.7741	-0.0107	0.4804

- 2. Discuss your findings. Do these baseline characteristics appear balanced? (1 point)
- 3. Calculate the treatment effect of receiving a voucher and the statistical significance for each of the outcomes, filling out the table below. Discuss your findings— What do you conclude about the effectiveness of the voucher program? (3 points)

	(1)	(2)	(3)	(4)
	Control	Treated	Difference	
Outcome	Mean	Mean	(2)- (1)	p-value
working_end	0.5162	0.5107	-0.0054	0.7658
happy_end	0.7125	0.7722	0.0496	0.0016

4. Calculate the treatment effect and statistical significance of receiving a voucher on working at endline, $working_end$, and happiness at endline, $happy_end$, separately for each city, filling out the table below. Discuss your results. (3 pts)

²Welch's two-sample t-test could also be used here, which does not assume equal variance. Since all the outcomes are binary, a Z-test could also be used here too. But there's a reason we want you to do it this way.







Winter 2023 Problem Set 1 PPHA 311

		(1)	(2)	(3)	(4)
		Control	Treated	Difference	
City	Outcome	Mean	Mean	(2)-(1)	p-value
Baltimore	working_end	0.6100	0.4897	-0.1202	0-0142
	happy_end	0.7798	0 8 077	0.0228	6.5665
Boston	working_end	0.5078	0.4787	-0.0290	0.4637
	happy_end	0.7382	0.7830	6.0447	0.1822
Chicago	working_end	0.5300	0.5184	-0.016	0.7889
	happy_end	0.7650	P9030	0.443	6.2031
Los Angeles	working_end	0.4694	0.5294	0.0509	0.1083
	happy_end	0.6977	0.7505	0.0528	D.1174
New York City	working_end	0.5130	0.5268	0.0137	6.7290
	happy_end	0.6652	0.7276	0.0624	0.0847

- 5. Have a look at the variable *moved_using_voucher_end*. Did everyone who received a voucher move? How does it affect the interpretation of our results if not everyone moves? How does the interpretation change if everyone had moved? (You don't need to do any calculations for full credit, only discussion is required) (1 pt)
- 6. As is common in experiments, everyone in the MTO study volunteered to participate. In what ways might this affect the external validity of the results? (1 pt)
- 7. Let's try redoing Part (3) using a simple linear regression.

For each of the two outcomes in Part (3) above, run a simple linear regression of the outcome on a treatment indicator:

$$Y_i = \beta_0 + \beta_1 Treated_i + u_i$$

Compare to your results in Part (3) above.

For this question, we recommend using the lm() regression function in R. (3 points)

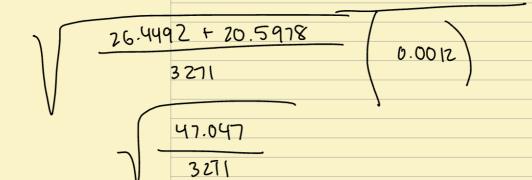






$$t = \frac{\overline{Y}_T - \overline{Y}_C}{\sqrt{\frac{(N_T - 1)s_T^2 + (N_C - 1)s_C^2}{N_T + N_C - 2}(\frac{1}{N_C} + \frac{1}{N_T})}}$$

$$\frac{-0.0057}{\sqrt{\frac{(2133)(0.0124)+(1138)(0.0181)}{2134+1139-2}} \left(\frac{1}{1139} + \frac{1}{2134}\right)$$



$$\sqrt{0.0143} = 0.1195 \quad (0.0012)$$