Assignment 3 Notebook

Prateek Naharia

Please read this code before starting. There are several ways to run regressions in R. I'd like us to use the function 'feols' with the option se = 'hetero'. We will discuss this in class.

```
# Code that shows how different regression functions work:
this_reg_lm <- lm(racism.scores.post.2mon ~ any_treatment,
                  data = tweets_data)
this_reg_lm
```

```
Call:
lm(formula = racism.scores.post.2mon ~ any_treatment, data = tweets_data)

Coefficients:
  (Intercept)  any_treatment
      0.25217       -0.08277
```

```
this_reg_feols <- feols(racism.scores.post.2mon ~ any_treatment,
                        data = tweets_data)
this_reg_feols
```

```
OLS estimation, Dep. Var.: racism.scores.post.2mon
Observations: 243
Standard-errors: IID
               Estimate Std. Error  t value   Pr(>|t|)
(Intercept)    0.252171   0.053558  4.70835 4.2164e-06 ***
any_treatment -0.082774   0.060411 -1.37019 1.7190e-01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.384622   Adj. R2: 0.003613
```

```r
this_reg_feols_robust <- feols(racism.scores.post.2mon ~ any_treatment,
                               data = tweets_data, se = 'hetero')
this_reg_feols_robust
```

```
OLS estimation, Dep. Var.: racism.scores.post.2mon
Observations: 243
Standard-errors: Heteroskedasticity-robust
               Estimate Std. Error  t value   Pr(>|t|)
(Intercept)    0.252171   0.063377  3.97891 9.1629e-05 ***
any_treatment -0.082774   0.068648 -1.20578 2.2908e-01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.384622   Adj. R2: 0.003613
```

```r
# this_reg_feols_inter <- feols(racism.scores.post.2mon ~ any_treatment*anonymity,
#                                data = tweets_data)

#racism.scores.post.2mon
# modelsummary(list(this_reg_lm, this_reg_feols, this_reg_feols_robust),
#              stars = TRUE, gof_omit = "AIC|BIC|Log|F|RMSE|Adj")
#
# modelsummary(list(this_reg_lm, this_reg_feols, this_reg_feols_robust),
#              coef_map = c('any_treatment' = 'Treatment',
#                'anonymity' = 'Anonymity',
#                'any_treatment:anonymity' =
#.                'Treatment * Anon.',
#                '(Intercept)' = 'Constant'),
#                stars = TRUE,
#                gof_omit = "AIC|BIC|Log|F|RMSE|Adj")
```

1.a: Of the above variables, please identify all that may be 'good' control variables in a regression where the outcome is 'racism.scores.post.2mon' and the regressor is 'any_treatment'?

Good control variables in a regression would be:- 'anonymity', 'log.followers', and 'racism.scores.pre.2mon'. Since these three are not affected by the treatment group Also, by controlling for these variables, we can isolate the impact of the treatment on the outcome of interest.

Note: By good control variables, I mean those which do not prevent a causal interpretation of the coefficient on treatment.

1.b: Run a regression of 'racism.scores.post.2mon' on 'any_treatment'.

What do we learn from this regression about the effect of the treatment? Please explain in words in addition to just returning the number.

Below is an example regression and how you should output it. In this regression, our outcome variable is 'anonymity' and our explanatory variable is 'log.followers'. This regression tells us whether there is correlation between anonymous twitter accounts and those who have a lot of followers (in our dataset).

```
reg_anon_followers <- feols(anonymity ~ log.followers, data = tweets_data, se = 'hetero')
etable(reg_anon_followers)
```

```
                reg_anon_follow..
Dependent Var.:         anonymity

Constant        1.451*** (0.1123)
log.followers     0.0176 (0.0181)
--------------- -----------------
S.E. type       Heteroskeda.-rob.
Observations                  243
R2                        0.00245
Adj. R2                  -0.00169
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Repeat the above exercise, but to answer question 1.b.

This regression models the relationship between the treatment received (any_treatment) and the level of racism (racism.scores.post.2mon) in the two months after the experiment.

Answer 1.b

```
reggn_rac_treat <- feols(racism.scores.post.2mon ~ any_treatment, data = tweets_data, se =
etable(reggn_rac_treat)
```

```
                      reggn_rac_treat
Dependent Var.: racism.scores.post.2mon

Constant             0.2522*** (0.0634)
any_treatment          -0.0828 (0.0687)
--------------- -----------------------
S.E. type       Heteroskedasticity-rob.
Observations                        243
R2                              0.00773
Adj. R2                         0.00361
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Observations are as:- 1. The coefficient is not statistically significant, as indicated - lack of a star(*) near the coefficient of any_treatment.

2. The coefficient of 'any_treatment' is -0.0828 and its standard error is 0.0687. The negative sign indicates that receiving any treatment is associated with a decrease in the racism score after 2 months.
3. R-Square value of 0.00773, gives us the idea that 'any_treatment' explains only a small proportion of the variation in the dependent variable 'racism.scores.post.2mon'.

Overall, based on this regression, we cannot draw a conclusion about the effect of treatment on the racism score after 2 months, as the coefficient is not statistically significant.

1.c Add the variables from a) as controls into the regression from b). What happens to our estimate of the effect of the treatment and its standard error? Why does this happen in words? YOUR ANSWER

```r
reg_all <- feols(racism.scores.post.2mon ~ any_treatment +
                   anonymity + log.followers +
                   racism.scores.pre.2mon,
                 data = tweets_data, se = 'hetero')
```

```
NOTE: 1 observation removed because of NA values (RHS: 1).
```

```r
etable(reg_all)
```

```
                                        reg_all
Dependent Var.:        racism.scores.post.2mon

Constant                        0.0568 (0.1002)
any_treatment                   0.0221 (0.0457)
anonymity                       0.0181 (0.0225)
log.followers                  -0.0014 (0.0163)
racism.scores.pre.2mon      0.7659*** (0.1287)
______________________ ______________________
S.E. type              Heteroskedasticity-rob.
Observations                               242
R2                                     0.27703
Adj. R2                                0.26482
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Change observed in coefficient value -> 'any_treatment' from -0.0828 to 0.0221 -> Therefore, estimated effect of the treatment is +ve.
2. Standard error has been observed as reduced from 0.0687 to 0.0457, as this improves the accuracy or estimate of treatment effect has become more precise.
3. It happens, since 'anonymity','log.followers','racism.scores.pre.2mon'(covariates) are highly explanatory of 'racism.scores.post.2mon'.
4. Most favorable predictor for 'racism.scores.post.2mon' is 'racism.scores.pre.2mon' (racism.scores.pre.2mon 0.7659*** (0.1287)).

Overall, this regression states that racism.scores.pre.2mon is statistically significant & +ve correlated with the outcome (racism.scores.post.2mon)

1.d Use a regression to check for differences between the treatment and control for one of the variables identified in a). Also use the prop.test function to check whether the randomization proportion was intended. Based on these results, should we be concerned that the randomization was done improperly?

Hint: we can do this by comparing the pre-treatment outcomes between the treatment and control group. If there are significant differences, then there may be a problem with the experiment.

Note: Each arm of the experiment was assigned with equal probability and there are 4 treatment arms and one control.

To check for differences between the treatment and control group for one of the variables, we can use a t-test. We can compare the means of the variable in question between the two groups to see if there is a significant difference. For example, if we want to compare log.followers between the treatment and control groups, we can do the following:

```
reg_treatment <- feols(racism.scores.pre.2mon ~ any_treatment,
                       data = tweets_data,
                       se = 'hetero')
```

```
NOTE: 1 observation removed because of NA values (LHS: 1).
```

```
etable(reg_treatment)
```

```
                         reg_treatment
Dependent Var.: racism.scores.pre.2mon

Constant              0.2277** (0.0707)
any_treatment        -0.1349. (0.0713)
_______________ ______________________
S.E. type        Heteroskedastici.-rob.
Observations                       242
R2                             0.04342
Adj. R2                        0.03943
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(reg_treatment)
```

```
OLS estimation, Dep. Var.: racism.scores.pre.2mon
Observations: 242
Standard-errors: Heteroskedasticity-robust
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    0.227667   0.070694  3.22046 0.0014568 **
any_treatment -0.134883   0.071277 -1.89238 0.0596436 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.260037   Adj. R2: 0.039434
```

As we can see above, Dependent variable : racism.scores.pre.2mon Independent variable : any_treatment

In the above output, 1. The intercept has an estimated coefficient of 0.227667 with the standard error of 0.070694 & t value of 3.22046

2. P-Value associated with intercept is 0.0014568, which is lesser than 0.05 -> referencing as statistically significant.
3. The coefficient for the variable "any_treatment" is -0.1349, with a standard error of 0.0713, a t-value of -1.89238.
4. P-value associated with 'any_treatment' is 0.0596436, which is greater than 0.05. This suggests that there is no statistically significant evidence that "any_treatment" is related to "racism.scores.pre.2mon".

Prop.Test 1 Control = 0.2 4 Treatment arms = 0.8 5 equal distributions between control & treatment. Data Split - 0.8 & 0.1 split

```r
prop.test(sum(tweets_data$any_treatment == 1)
          ,nrow(tweets_data)
          , 0.8, alternative = "two.sided")
```

```
    1-sample proportions test with continuity correction

data:  sum(tweets_data$any_treatment == 1) out of nrow(tweets_data), null probability 0.8
X-squared = 0.21631, df = 1, p-value = 0.6419
alternative hypothesis: true p is not equal to 0.8
95 percent confidence interval:
 0.7280013 0.8347620
sample estimates:
        p
0.7860082
```

Above observations as:-

1. In this case, the p-value of 0.6419 is not less than the significance level (usually 0.05), which means that there is not enough evidence to reject the null hypothesis.
2. In other words, the proportion of any_treatment in the sample (0.7860082) is not significantly different from 0.8.
3. The 95% CI Interval [0.7280013, 0.8347620], means that the true proportion of any_treatment in the population lies between 0.7280013 and 0.8347620.

After obtaining the results, we should not be concerned that the randomization was done improperly.

1.e We would like to know whether treatment arm 2 or treatment arm 3 is statistically significantly better at reducing racist behavior. Perform a t.test or regression and test for the null hypothesis that treatment arm 2 has the same effect as treatment arm 3.

```
treat_arm2 <- tweets_data[treatment_arm == 2, racism.scores.post.2mon]
treat_arm3 <- tweets_data[treatment_arm == 3, racism.scores.post.2mon]

t.test(treat_arm2, treat_arm3)
```

```
	Welch Two Sample t-test

data:  treat_arm2 and treat_arm3
t = 1.6246, df = 64.539, p-value = 0.1091
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0151519  0.1472047
sample estimates:
 mean of x  mean of y
0.13892962 0.07290323
```

Following observations:-

The t-statistic is 1.6246, which indicates the difference between the sample means. The degrees of freedom is calculated as 64.539. The p-value of 0.1091 not low enough to reject the null hypothesis, hence cannot reject the null hypothesis.

Therefore, we cannot conclude that there is a significant difference in the racism.scores.post.2mon scores between those who received treatment arm 2 and those who received treatment arm 3.

95% CI is [-0.015 to 0.147] - true difference in means falls in within this range, contains 0

This means that there is not enough evidence to conclude that treatment arm 2 is statistically significantly better at reducing racist behavior than treatment arm 3 OR Cannot say that treatment arm 3 is statistically significantly better/worse at reducing racist behavior.

2.a Describe the treatment in the first experiment and the unit of randomization. What share was randomized to the treatment?

(This refers to the experiment conducted in August 2015, the first experiment described in the introduction of the paper.)

In the first experiment of the StubHub Price Salience study, the treatment group was exposed to a back-end fee purchasing experience. In this experience, the mandatory fees were displayed only after the consumers had chosen a specific ticket and moved on to the checkout page. On the contrary, the control group was exposed to an upfront fee (UF) experience. The unit of randomization was the individual user visiting the website, at user level with half (50 %) share of the U.S. users being randomly selected for the treatment group, and the other 50% for the control group.

2.b Table II displays a randomization / balance check. A randomization check is a regression where the dependent variable occurs before the experiment. It should be very unlikely that there are substantial differences in before experiment variables if the experiment was done properly. Suggest a variable not used by the authors that would be appropriate to include in a balance check.

Variable appropriate to include in a balance check:-

The number of transactions by user before the experiment or considering user demographics & purchase behavior, it would be appropriate to also include browsing behavior. To further verify that the treatment & control groups are well-balanced. This information can provide additional insight into the potential impact of the Back-end fees purchasing experience on the users' purchasing behavior.

2.c What is the effect of the treatment on the Propensity to Purchase at least one product? Calculate the 95% confidence interval for this estimate.

Considering Table III

Propensity to Purchase at Least Once, treatment effect = 14.1%

Formula -> 95% CI as:- CI = (ATE_hat(Estimate Avg Treatment Effect) - 1.96 * SE(Standard Error), ATE_hat + 1.96 * SE)

```r
CI_U <- (0.141 + 1.96 * 0.09)
CI_L <- (0.141 - 1.96 * 0.09)
print(paste0('95% Confidence Interval Upper Limit: ',CI_U))
```

```
[1] "95% Confidence Interval Upper Limit: 0.3174"
```

```r
print(paste0('95% Confidence Interval Lower Limit: ',CI_L))
```

```
[1] "95% Confidence Interval Lower Limit: -0.0354"
```

```r
cat('95% Confidence Interval',CI_L,',' , CI_U)
```

```
95% Confidence Interval -0.0354 , 0.3174
```

Also, CI=(0.141 - 1.96 * 0.09, 0.141 + 1.96 * 0.09) CI=(-0.0354, 0.3174) 14.1% is 0.141.

2.d Suppose the authors randomized by city of the event. Name one benefit that may occur as a result of this randomization strategy and one harm.

Benefit of randomizing at the city level:- Randomizing by city of the event could result in one benefit, which is that all users in a specific city would have the same purchasing experience. This could lead to a smoother purchase process for users who are booking multiple events within the same city.

One Harm:- The tendency of the Back-end Fees (BF) strategy to drive more sales in cities with events could result in lower traffic to events in cities with the Upfront Fees (UF) strategy, potentially violating the non-interference assumption. Additionally, the variability in event venues and types within cities could result in differences in the results, making it difficult to generalize the findings to other cities or geographic regions.

2.e Suppose that you are the product manager for the monetization team at Stubhub. Based on the evidence presented above, would you launch the treatment to the entire site? The answer should be 1 paragraph. It should consist of an answer (Yes, no), and two pieces of evidence relating to that recommendation. Case participation will also constitute part of this grade.

By looking at the "Table III: Effect of Salience on Purchasing", considering "Cookie 10-day Revenue", with 20.64% rise. (TREATMENT), "Propensity to Purchase at Least Once", with 14.1 rise. Being a product manager YES, I would launch the treatment to the entire.

How long did this assignment take you to do (hours)? How hard was it (easy, reasonable, hard, too hard)?

2 days, hard