Problem Set 3

Experiments and Causality

```
# load packages
library(data.table)
library(foreign)
```

0. Write Functions

You're going to be doing a few things a *number* of times – calculating robust standard errors, calculating clustered standard errors, and then calculating the confidence intervals that are built off these standard errors.

After you've worked through a few of these questions, I suspect you will see places to write a function that will do this work for you. Include those functions here, if you write them.

```
# import the libraries used
library(sandwich)
library(lmtest)
library(stargazer)
```

```
get_robust_se <- function(mod) {
  sqrt(diag(vcovHC(mod)))
}
```

1. Replicate Results

Skim Broockman and Green's paper on the effects of Facebook ads and download an anonymized version of the data for Facebook users only.

```
d <- fread("./data/broockman_green_anon_pooled_fb_users_only.csv")
```

While running the regressions, we realized that there are some records having NA values for the dependent variables. We will take a look at them and decide what to do here.

```
# get an overview of the dataset
summary(d)
```

```
##      studyno         treat_ad         cluster            name_recall
##   Min.   :1.000   Min.   :0.0000   Length:2706        Min.   :0.0000
##   1st Qu.:1.000   1st Qu.:0.0000   Class :character   1st Qu.:0.0000
##   Median :1.000   Median :0.0000   Mode  :character   Median :0.0000
##   Mean   :1.496   Mean   :0.4213                      Mean   :0.3887
##   3rd Qu.:2.000   3rd Qu.:1.0000                      3rd Qu.:1.0000
##   Max.   :2.000   Max.   :1.0000                      Max.   :1.0000
##                                                       NA's   :5
##   positive_impression
##   Min.   :0.0000
##   1st Qu.:0.0000
##   Median :0.0000
##   Mean   :0.2603
```

```
##  3rd Qu.:1.0000
##  Max.   :1.0000
##  NA's   :5
```

```
# look at those records having NAs for dependent variables
d[is.na(name_recall) | is.na(positive_impression), ]
```

```
##    studyno treat_ad                         cluster name_recall
## 1:       2        0 Study 2, Cluster Number 229          NA
## 2:       2        0 Study 2, Cluster Number 287          NA
## 3:       2        0 Study 2, Cluster Number 374          NA
## 4:       2        0 Study 2, Cluster Number 501          NA
## 5:       2        0 Study 2, Cluster Number 539          NA
##    positive_impression
## 1:                  NA
## 2:                  NA
## 3:                  NA
## 4:                  NA
## 5:                  NA
```

```
# the records having NAs for name_recall are the same as those having NAs for positive_impression. Sinc
```

```
d <- d[!is.na(name_recall), ]
```

a. Using regression without clustered standard errors (that is, ignoring the clustered assignment), compute a confidence interval for the effect of the ad on candidate name recognition in Study 1 only (the dependent variable is "name_recall").

- **Note**: Ignore the blocking the article mentions throughout this problem.
- **Note**: You will estimate something different than is reported in the study.

```
mod_study_1 <- d[studyno == 1, lm(name_recall ~ treat_ad)]
summary(mod_study_1)
```

```
##
## Call:
## lm(formula = name_recall ~ treat_ad)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1825 -0.1825 -0.1727 -0.1727  0.8273
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.182469   0.016142  11.304   <2e-16 ***
## treat_ad    -0.009798   0.021012  -0.466    0.641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3817 on 1362 degrees of freedom
## Multiple R-squared:  0.0001596,  Adjusted R-squared:  -0.0005745
## F-statistic: 0.2174 on 1 and 1362 DF,  p-value: 0.6411
```

```
# print(summary(mod_study_1)$coefficient[2,1] + 2*summary(mod_study_1)$coefficient[2,2])
```

The 95 % confidence interval for the effect of the ad on candidate name recognition in study 1 is (-0.0518223, 0.0322265)

2

b. What are the clusters in Broockman and Green's study? Why might taking clustering into account increase the standard errors? **The clusters in the study were formed by the combination of age range, gender and location. For example, a cluster could be 30-31, San Francisco, male. Clustered treatment assignment with clusters that share more in common within the groups than across the groups will likely increase the variance of our estimate by chance because it is more likely that subjects in treatment have different potential outcomes from subjects in control. In this study, since age, gender and location are all correlated with one's political views and knowledge, so it's reasonable to expect that subjects in different clusters are more different than similar in expectation. Therefore, we would expect the standard error to increase.**

c. Now estimate a regression that estimates the effect of the ad on candidate name recognition in Study 1, but this time take take clustering into account. (hint: The estimation of the *model* does not change, only the estimation of the standard errors.) If you're not familiar with how to calculate these clustered and robust estimates, there is a demo worksheet that is available in our course repository: `./week_05/cluster_and_robust.Rmd`.

```
mod_study_1$vcovCL_ <- vcovCL(mod_study_1, cluster = d[studyno == 1, cluster])
coeftest(mod_study_1, mod_study_1$vcovCL_)


##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1824687  0.0184915  9.8677    <2e-16 ***
## treat_ad    -0.0097979  0.0237536 -0.4125    0.6801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# print(coeftest(mod_study_1, mod_study_1$vcovCL_)[2,2])
```

The 95 % confidence interval for the effect of the ad on candidate name recognition in study 1 when accounting for clustering is (-0.0573051, 0.0377094)

d. Again, run a regression to test for the effect of the ad on candidate name recognition using clustered standard errors, but this time conduct it only for Study 2. How can you employ some form of slicing to make the code you've written in parts (c) and (d) very similar?

```
mod_study_2 <- d[studyno == 2, lm(name_recall ~ treat_ad)]
# summary(mod_study_2)
mod_study_2$vcovCL_ <- vcovCL(mod_study_2, cluster = d[studyno == 2, cluster])
coeftest(mod_study_2, mod_study_2$vcovCL_)


##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6057884  0.0181889 33.305     <2e-16 ***
## treat_ad    -0.0028033  0.0355033 -0.079     0.9371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e. Run a regression to test for the effect of the ad on candidate name recognition, but this time the entire sample from both studies. Do not take into account which study the data is from (more on this in a moment), but just pool the data. What is the treatment effect estimate? What is the p-value associated with this treatment effect test?

3

```r
mod_study_all_1 <- d[, lm(name_recall ~ treat_ad)]
summary(mod_study_all_1)
```

```
##
## Call:
## lm(formula = name_recall ~ treat_ad)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4542 -0.4542 -0.2991  0.5458  0.7009
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.45420    0.01219  37.262   <2e-16 ***
## treat_ad    -0.15507    0.01876  -8.265   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4816 on 2699 degrees of freedom
## Multiple R-squared:  0.02469,    Adjusted R-squared:  0.02432
## F-statistic: 68.31 on 1 and 2699 DF,  p-value: < 2.2e-16
```

```r
mod_study_all_1$vcovCL_ <- vcovCL(mod_study_all_1, cluster = d[, cluster])
coeftest(mod_study_all_1, mod_study_all_1$vcovCL_)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  0.454196   0.018576 24.4504 < 2.2e-16 ***
## treat_ad    -0.155073   0.026730 -5.8014 7.344e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# coeftest(mod_study_all_1, mod_study_all_1$vcovCL_)[2,1]
# coeftest(mod_study_all_1, mod_study_all_1$vcovCL_)[2,4]
```

We see that when the clustering is taken into consideration, and the data from the two studies are pooled together, we have an estimated treatment effect of -0.1550732 with the p-value being 7.344e-09

f. Now, repeat the last question-part, but this time include a variable that identifies whether an observation was generated during Study 1 or Study 2. What is estimated in the "Study 2 Fixed Effect"? What is the treatment effect estimate and associated p-value? Think a little bit more about the treatment effect that you've estimated: can this treatment effect be *different* between Study 1 and Study 2? Why or why not?

```r
d <- d[, study2 := studyno - 1]
mod_study_all_2 <- d[, lm(name_recall ~ treat_ad + study2)]
# summary(mod_study_all_2)
mod_study_all_2$vcovCL_ <- vcovCL(mod_study_all_2, cluster = d[, cluster])
coeftest(mod_study_all_2, mod_study_all_2$vcovCL_)
```

```
##
## t test of coefficients:
##
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1806848  0.0169702 10.6472   <2e-16 ***
## treat_ad    -0.0067752  0.0204154 -0.3319     0.74
## study2       0.4260988  0.0206970 20.5875   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimated study 2 fixed effect is 0.42 with a p-value smaller than 2e-16. However, the fixed effect only suggests that the study two for some reason generates higher rates of name recall which is caused by factors specific to study 2 but not the treatment. I don't think the treatment effects are different between the two studies. Based on what we've estimated in part (c) and (d), we see that when the two studies' data are analyzed separately, both results suggested that we have failed to reject the null hypothesis that the treatment effect is zero. Therefore, we do not have any evidence that suggests the treatment effects for the two studies are different.

g. Conduct a formal test – it must have a p-value associated with the test – for whether the treatment effects are different in Study 1 than Study 2. If they are different, why do you suppose they differ? Is one of the results "biased"? Why or why not? (Hint: see pages 75-76 of Gerber and Green, with more detailed discussion optionally available on pages 116-121.)

```
mod_study_all_3 <- d[, lm(name_recall ~ treat_ad + study2 + treat_ad * study2)]
# summary(mod_study_all_3)
mod_study_all_3$vcovCL_ <- vcovCL(mod_study_all_3, cluster = d[, cluster])
coeftest(mod_study_all_3, mod_study_all_3$vcovCL_)
```

```
##
## t test of coefficients:
##
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.1824687  0.0184880  9.8696   <2e-16 ***
## treat_ad        -0.0097979  0.0237491 -0.4126   0.6800
## study2           0.4233197  0.0259296 16.3257   <2e-16 ***
## treat_ad:study2  0.0069945  0.0427010  0.1638   0.8699
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
study2_treat <- round(d[study2 == 1, mean(treat_ad)] * 100, 1)
study1_treat <- round(d[study2 == 0, mean(treat_ad)] * 100, 1)
```

We ran a regression with the data of the two studies together by including the treatment, study number and the interaction between the two variables as the terms on the right hand side of the equation. This way we can formally test the heterogeneous treatment effects of the two study by looking at the significance of the coefficient of the interaction term. Since the test suggests that there is a 87% chance that we would get a coefficient estimate of the interaction term at least as large as our estimate under the null hypothesis, we failed to reject the null (that the treatment effects are different between the two studies).

Looking at the results from the regression ran from part (a) to (f), I notice that the result from part (e) is a biased estimate that suggests a non-zero treatment effect. Since the two studies were run separately, when looking at the two studies together, they effectively form "a larger study that is blocked on the study number". Yet, since the two studies are assigning different proportion of the subjects to treatment (study 2 has 25.1% in treatment and study 1 has 59% in treatment), a naive strategy of pooling the results together without adjusting for the different probability of treatment administration generates a biased estimate. This adjustment can be done by controlling for study fixed effect as we

did in part (f) and (g).

2. Peruvian Recycling

Look at this article about encouraging recycling in Peru. The paper contains two experiments, a "participation study" and a "participation intensity study." In this problem, we will focus on the latter study, whose results are contained in Table 4 in this problem. You will need to read the relevant section of the paper (starting on page 20 of the manuscript) in order to understand the experimental design and variables. (*Note that "indicator variable" is a synonym for "dummy variable," in case you haven't seen this language before.*)

a. In Column 3 of Table 4A, what is the estimated ATE of providing a recycling bin on the average weight of recyclables turned in per household per week, during the six-week treatment period? Provide a 95% confidence interval. **The 95% confidence interval is (0.123, 0.251)**

b. In Column 3 of Table 4A, what is the estimated ATE of sending a text message reminder on the average weight of recyclables turned in per household per week? Provide a 95% confidence interval. **The 95% confidence interval is (-0.102, 0.054)**

c. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of providing a recycling bin? **Percentage of visits turned in bag, Avg. no. of bins turned in per week, Avg. weight (in kg) of recyclables turned in per week, and Avg. market value of recyclables given per week.**

d. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of sending text messages? **None**

e. Suppose that, during the two weeks before treatment, household A turns in 2kg per week more recyclables than household B does, and suppose that both households are otherwise identical (including being in the same treatment group). From the model, how much more recycling do we predict household A to have than household B, per week, during the six weeks of treatment? Provide only a point estimate, as the confidence interval would be a bit complicated. This question is designed to test your understanding of slope coefficients in regression. **We would expect the household A to turn in 0.562 kg of recyclables more**

f. Suppose that the variable "percentage of visits turned in bag, baseline" had been left out of the regression reported in Column 1. What would you expect to happen to the results on providing a recycling bin? Would you expect an increase or decrease in the estimated ATE? Would you expect an increase or decrease in the standard error? Explain your reasoning. **I would expect the estimated ATE to have no change and the standard error to increase. Since whether to receive a recycling bin is randomly assigned, it is independent of the baseline number. Therefore, knowing how much a household recycled pre-treatment should not give us any information on wheter the household receives a bin as well as the expected effect of receiving bins on recyclables turned in post-treatment. However, we do expect the standard error to rise. Including pre-treatment measure could help explain the variability in our data. With less variability (uncertainty) to explain by the treatment, the the standard error of the ATE estimate shrinks; therefore, when we exclude pre-treatment measure, we expect to see the standard error of our estimate to increase.**

g. In column 1 of Table 4A, would you say the variable "has cell phone" is a bad control? Explain your reasoning. **I don't think it is a bad control. Because whether or not a subject unit has a cell phone is not an outcome affected by the treatment in anyway. Therefore, including "has cell phone" will not bias the estimate.**

h. If we were to remove the "has cell phone" variable from the regression, what would you expect to happen to the coefficient on "Any SMS message"? Would it go up or down? Explain your reasoning. **Since having a cell phone is positively correlated with receiving SMS messages (one would need**

a phone to be in the SMS group) and also positively correlated with some of the treatment outcomes (for column 2-4), when having a cell phone is omitted, SMS messages could **"take its credit". Therefore, we would expect to see the "Any SMS message" coefficient estimate to go up in those columns when having a cell is removed from the equation.**

3. Multifactor Experiments

Staying with the same experiment, now lets think about multifactor experiments.

a. What is the full experimental design for this experiment? Tell us the dimensions, such as 2x2x3. (Hint: the full results appear in Panel 4B.) **It is 4x3. The study randomizes the treatment by looking at whether a household has a cell phone or not, and then randomly assigning people who have a cell phone to receive personal SMS, generic SMS or no SMS. Therefore, we end up having 4 different dimensions in the first part, which are no phone (no message), having phone + no message, having phone + personal message and having phone + generic message. Then the second part is randomized by blocking on the message groups; therefore, each variation in the first part is associated with the three different treatment variations related to bins (bin with sticker, bin without stick, no bin). So in the end, we have a 4x3 design with 12 different treatment groups.**

b. In the results of Table 4B, describe the baseline category. That is, in English, how would you describe the attributes of the group of people for whom all dummy variables are equal to zero? **The baseline cateory is the group that didn't receive any bin, SMS messgae and has no phone**

c. In column (1) of Table 4B, interpret the magnitude of the coefficient on "bin without sticker." What does it mean? **It means we expect the chance that a household will remit some recyclables on a given visit to increase by 3.5 percentage point.**

d. In column (1) of Table 4B, which seems to have a stronger treatment effect, the recycling bin with message sticker, or the recycling bin without sticker? How large is the magnitude of the estimated difference? **Recycling bin with sticker seems to have a stronger effect. The magnitude is 2 percentage point**

e. Is this difference you just described statistically significant? Explain which piece of information in the table allows you to answer this question. **In the row, F-test p-value (1) = (2), we have the p-value from the F-test that tests whether the difference between the two coefficients is statistically significant. Since the p-value is 0.31, I don't think there is evidence that suggests the difference in (d) is statistically significant.**

f. Notice that Table 4C is described as results from "fully saturated" models. What does this mean? Looking at the list of variables in the table, explain in what sense the model is "saturated." **Fully saturated model means there is a term associated with each of the treatment variation. As calculated in part (a), there are 12 treatment combinations in the study. In table 4C, the first eleven terms (from Generic SMS message + Bin with sticker(1) to No SMS message + No Bin) each represents a treatment combination. The omitted category, "No phone + No Bin" is represented by one of the coefficients omitted from the table with the street fixed effects.**

4. Now! Do it with data

Download the data set for the recycling study in the previous problem, obtained from the authors. We'll be focusing on the outcome variable Y="number of bins turned in per week" (avg_bins_treat).

```r
d <- read.dta("./data/karlan_data_subset_for_class.dta")
d <- data.table(d)
head(d)
```

```
##    street havecell avg_bins_treat base_avg_bins_treat bin sms bin_s bin_g
## 1:      7        1      1.0416666               0.750   1   1     1     0
## 2:      7        1      0.0000000               0.000   0   1     0     0
## 3:      7        1      0.7500000               0.500   0   0     0     0
## 4:      7        1      0.5416667               0.500   0   0     0     0
## 5:      6        1      0.9583333               0.375   1   0     0     1
## 6:      8        0      0.2083333               0.000   1   0     0     1
##    sms_p sms_g
## 1:     0     1
## 2:     1     0
## 3:     0     0
## 4:     0     0
## 5:     0     0
## 6:     0     0
```

```
## Do some quick exploratory data analysis with this data. There are some values in this data that seem
```

```r
# EDA
summary(d)
```

```
##      street            havecell        avg_bins_treat   base_avg_bins_treat
##  Min.   :-999.00   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:  69.00   1st Qu.:0.0000   1st Qu.:0.4167   1st Qu.:0.3750
##  Median : 131.50   Median :1.0000   Median :0.6250   Median :0.6250
##  Mean   :  68.81   Mean   :0.5908   Mean   :0.6811   Mean   :0.7363
##  3rd Qu.: 215.00   3rd Qu.:1.0000   3rd Qu.:0.8333   3rd Qu.:1.0000
##  Max.   : 263.00   Max.   :1.0000   Max.   :4.1667   Max.   :6.3750
##  NA's   :3         NA's   :1
##       bin               sms             bin_s             bin_g
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.3378   Mean   :0.3087   Mean   :0.1681   Mean   :0.1697
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##      sms_p             sms_g
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000
##  Mean   :0.1557   Mean   :0.1529
##  3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000
##
```

High-level overview of the column summary: We can see that there are some observations having NAs in columns, street or havecell. Since it's a small number (only 4 observations) compared to the total number of observation, 1785, we will drop those observations to keep the number of examples used in different models consistent.

Another decision that was made is that despite seeing some of the streets are labelled -999, they are still included because we are going to use the street variable as categorical variable.

Therefore, the value of the number doesn't matter so long as they are not missing. We suspect that those observations associated with -999 are the households with street unknown. By keeping -999, we will be predicting a fixed effect for those households associated with street -999.

```
d_clean = d[!is.na(street) & !is.na(havecell)]
```

a. For simplicity, let's start by measuring the effect of providing a recycling bin, ignoring the SMS message treatment (and ignoring whether there was a sticker on the bin or not). Run a regression of Y on only the bin treatment dummy, so you estimate a simple difference in means. Provide a 95% confidence interval for the treatment effect.

```
mod_a <- d_clean[, lm(avg_bins_treat ~ bin)]
summary(mod_a)
```

```
##
## Call:
## lm(formula = avg_bins_treat ~ bin)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7690 -0.2609 -0.0526  0.1894  3.5308
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.63589    0.01175  54.124  < 2e-16 ***
## bin           0.13307    0.02024   6.574 6.42e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4038 on 1779 degrees of freedom
## Multiple R-squared:  0.02372,    Adjusted R-squared:  0.02317
## F-statistic: 43.22 on 1 and 1779 DF,  p-value: 6.415e-11
```

```
summary(mod_a)$coefficients[2,1] - 2 * summary(mod_a)$coefficients[2,2]
```

```
## [1] 0.09258711
```

The confidence interval is (0.0925871, 0.1735542)

b. Now add the pre-treatment value of Y as a covariate. Provide a 95% confidence interval for the treatment effect. Explain how and why this confidence interval differs from the previous one.

```
mod_b <- d_clean[, lm(avg_bins_treat ~ bin + base_avg_bins_treat)]
stargazer(
  mod_a,
  mod_b,
  type = 'text',
  add.lines = list(
    c('Street fixed effects', 'No', 'No')),
  model.numbers=FALSE,
  column.labels = c("Model (a)", "Model (b)")
)
```

```
##
## ========================================================================
##                                Dependent variable:
##                    -------------------------------------------------
```

```
##                                      avg_bins_treat
##                              Model (a)                   Model (b)
## -------------------------------------------------------------------
## bin                          0.133***                    0.124***
##                              (0.020)                     (0.017)
##
## base_avg_bins_treat                                      0.390***
##                                                          (0.013)
##
## Constant                     0.636***                    0.352***
##                              (0.012)                     (0.014)
##
## -------------------------------------------------------------------
## Street fixed effects            No                          No
## Observations                  1,781                       1,781
## R2                            0.024                       0.338
## Adjusted R2                   0.023                       0.337
## Residual Std. Error     0.404 (df = 1779)         0.333 (df = 1778)
## F Statistic           43.218*** (df = 1; 1779) 453.012*** (df = 2; 1778)
## =================================================================
## Note:                                      *p<0.1; **p<0.05; ***p<0.01
```

The confidence interval is (0.0908772, 0.1576021). Since we are including the baseline which is predictive of the observed outcome, it helps to explain some of the variance in the outcome and shrink the standard error of the treatment effect estimate (the width of the confidence interval went from 0.0809671 to 0.0667249). By the way, we also could see some change in the point estimate, that could have been caused by the trivial correlation between the pre-treatment baseline and treatment assignment.

c. Now add the street fixed effects. (You'll need to use the R command factor().) Provide a 95% confidence interval for the treatment effect.

```r
mod_c <- d_clean[, lm(avg_bins_treat ~ bin + base_avg_bins_treat + as.factor(street))]
# summary(mod_3)
stargazer(
  mod_b,
  mod_c,
  type = 'text',
  omit = 'street',
  add.lines = list(
    c('Street fixed effects', 'No', 'Yes')
  ),
  model.numbers=FALSE,
  column.labels = c("Model (b)", "Model (c)")
)
```

```
##
## =================================================================
##                              Dependent variable:
##                   -----------------------------------------------
##                              avg_bins_treat
##                       Model (b)                   Model (c)
## -------------------------------------------------------------------
## bin                   0.124***                    0.114***
##                       (0.017)                     (0.017)
##
```

```
## base_avg_bins_treat              0.390***                    0.374***
##                                   (0.013)                     (0.014)
##
## Constant                         0.352***                    0.368***
##                                   (0.014)                     (0.032)
##
## -------------------------------------------------------------------------
## Street fixed effects                No                          Yes
## Observations                      1,781                       1,781
## R2                                0.338                       0.436
## Adjusted R2                       0.337                       0.372
## Residual Std. Error       0.333 (df = 1778)           0.324 (df = 1599)
## F Statistic          453.012*** (df = 2; 1778) 6.833*** (df = 181; 1599)
## =========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

The confidence interval is (0.0795512, 0.1478669).

d. Recall that the authors described their experiment as "stratified at the street level," which is a synonym for blocking by street. Explain why the confidence interval with fixed effects does not differ much from the previous one. **When blocking on a variable, the intention is to randomize within blocks so different blocks are more evenly represented in different treatment and control groups. Blocked design is aimed at minimizing the variance in the outcome variable that could be caused by the variables that we block on. Since most of the variance that could be caused by the `street` variable is already accounted for by blocking, we didn't see much change in the standard error of bin ATE when including street fixed effect into regression.**

e. Perhaps having a cell phone helps explain the level of recycling behavior. Instead of "has cell phone," we find it easier to interpret the coefficient if we define the variable " no cell phone." Give the R command to define this new variable, which equals one minus the "has cell phone" variable in the authors' data set. Use "no cell phone" instead of "has cell phone" in subsequent regressions with this dataset.

```
d_clean <- d_clean[, no_cell_phone := 1 - havecell]
```

f. Now add "no cell phone" as a covariate to the previous regression. Provide a 95% confidence interval for the treatment effect. Explain why this confidence interval does not differ much from the previous one.

```r
mod_f <- d_clean[, lm(avg_bins_treat ~ bin
                                  + base_avg_bins_treat
                                  + no_cell_phone
                                  + as.factor(street)
                  )
        ]
stargazer(
  mod_c,
  mod_f,
  type = 'text',
  omit = 'street',
  add.lines = list(
    c('Street fixed effects', 'No', 'Yes')
  ),
  model.numbers=FALSE,
  column.labels = c("Model (c)", "Model (f)")
)

##
## =========================================================================
```

```
## Dependent variable:
## ----------------------------------------------------
## avg_bins_treat
## Model (c) Model (f)
## ----------------------------------------------------
## bin 0.114*** 0.115***
## (0.017) (0.017)
##
## base_avg_bins_treat 0.374*** 0.373***
## (0.014) (0.014)
##
## no_cell_phone -0.050***
## (0.017)
##
## Constant 0.368*** 0.387***
## (0.032) (0.032)
##
## ----------------------------------------------------
## Street fixed effects No Yes
## Observations 1,781 1,781
## R2 0.436 0.439
## Adjusted R2 0.372 0.375
## Residual Std. Error 0.324 (df = 1599) 0.323 (df = 1598)
## F Statistic 6.833*** (df = 181; 1599) 6.875*** (df = 182; 1598)
## ====================================================================
## Note: *p<0.1; **p<0.05; ***p<0.01
```

The confidence interval is (0.0810108, 0.1491907). Since bin treatment is randomly assigned and looking at Table 1, Panel B in the paper, we see that the result of the random assignment for this study produced comparable proportion of bin treatment in the have-cell-phone group and the no-cell-phone group. Therefore, the variable cell-phone is expected to have little to no correlation with treatment. That is why when it is added to the model, we see little to no effect on our treatment's point estimation. Looking at the R2, we see that the inclusion of `no_cell_phone` barely explains more variance in the outcome. I think that is the reason why we also didn't see any change in the standard error of the effect of bin.

g. Now let's add in the SMS treatment. Re-run the previous regression with "any SMS" included. You should get the same results as in Table 4A. Provide a 95% confidence interval for the treatment effect of the recycling bin. Explain why this confidence interval does not differ much from the previous one.

```r
mod_g <- d_clean[, lm(avg_bins_treat ~ bin
                                      + base_avg_bins_treat
                                      + no_cell_phone
                                      + sms
                                      + as.factor(street)
                      )
              ]
stargazer(
  mod_f,
  mod_g,
  type = 'text',
  omit = 'street',
  add.lines = list(
    c('Street fixed effects', 'No', 'Yes')
```

```
  ),
  model.numbers=FALSE,
  column.labels = c("Model (f)", "Model (g)")
)
```

```
##
## ============================================================================
##                                Dependent variable:
##                         ----------------------------------------------------
##                                      avg_bins_treat
##                               Model (f)                   Model (g)
## -------------------------------------------------------------------------
## bin                          0.115***                    0.115***
##                               (0.017)                     (0.017)
##
## base_avg_bins_treat          0.373***                    0.373***
##                               (0.014)                     (0.014)
##
## no_cell_phone                -0.050***                   -0.047**
##                               (0.017)                     (0.020)
##
## sms                                                        0.005
##                                                           (0.021)
##
## Constant                     0.387***                    0.385***
##                               (0.032)                     (0.034)
##
## -------------------------------------------------------------------------
## Street fixed effects            No                          Yes
## Observations                  1,781                       1,781
## R2                            0.439                       0.439
## Adjusted R2                   0.375                       0.375
## Residual Std. Error     0.323 (df = 1598)           0.323 (df = 1597)
## F Statistic          6.875*** (df = 182; 1598) 6.834*** (df = 183; 1597)
## ============================================================================
## Note:                                      *p<0.1; **p<0.05; ***p<0.01
```

The confidence interval is (0.0809516, 0.1491557). Since bin treatment and sms treatment are assigned independent from each other and looking at Table 1, Panel B, we see that the result of the random assignment for this study produced comparable proportion of bin treatment in the sms group and the no-sms group. Therefore, the inclusion of sms should not have any effect on the point estimation of bin effect. Plus, since sms is a poor predictor of the outcome (its coefficient isn't statistically significant), the inclusion of sms also didn't have any impact on the standard error of bin effect.

h. Now reproduce the results of column 2 in Table 4B, estimating separate treatment effects for the two types of SMS treatments and the two types of recycling-bin treatments. Provide a 95% confidence interval for the effect of the unadorned recycling bin. Explain how your answer differs from that in part (g), and explain why you think it differs.

```
mod_h <- d_clean[, lm(avg_bins_treat ~ bin_s
                                        + bin_g
                                        + sms_p
                                        + sms_g
                                        + no_cell_phone
```

```
                                      + base_avg_bins_treat
                                      + as.factor(street)
                    )
            ]
stargazer(
  mod_h,
  type = 'text',
  omit = 'street',
  add.lines = list(
    c('Street fixed effects', 'Yes')
  ),
  model.numbers=FALSE,
  column.labels = c("Model (h)")
)
```

```
##
## ================================================
##                      Dependent variable:
##                  ----------------------------
##                        avg_bins_treat
##                          Model (h)
## ------------------------------------------------
## bin_s                      0.128***
##                            (0.022)
##
## bin_g                      0.103***
##                            (0.022)
##
## sms_p                      -0.008
##                            (0.025)
##
## sms_g                       0.020
##                            (0.025)
##
## no_cell_phone              -0.046**
##                            (0.020)
##
## base_avg_bins_treat        0.374***
##                            (0.014)
##
## Constant                   0.385***
##                            (0.034)
##
## ------------------------------------------------
## Street fixed effects          Yes
## Observations                 1,781
## R2                           0.440
## Adjusted R2                  0.375
## Residual Std. Error    0.323 (df = 1595)
## F Statistic         6.769*** (df = 185; 1595)
## ================================================
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

The confidence interval is (0.0594125, 0.1469679). The point estimate of treament effect

of generic bins is lower and the standard error higher compared with those of bins in general. I think the higher standard error is a reflection of the smaller sample size since now we are at looking only a subset of the group we examined in part (g). I also noticed that bin in general (part g) has the ATE that is approximately the mean of the ATEs of bin with sticker and bin without sticker. I think this could be expained by the fact that subjects are pretty evenly distributed between the treatments (i.e. the number of subjects receiving bins with sticker is about the same as the number of subjects receiving bins without sticker).

5. A Final Practice Problem

Now for a fictional scenario. An emergency two-week randomized controlled trial of the experimental drug ZMapp is conducted to treat Ebola. (The control represents the usual standard of care for patients identified with Ebola, while the treatment is the usual standard of care plus the drug.)

Here are the (fake) data.

```
d <- fread("./data/ebola_rct2.csv")
head(d)
```

```
##     temperature_day0 vomiting_day0 treat_zmapp temperature_day14
## 1:          99.53168             1           0          98.62634
## 2:          97.37372             0           0          98.03251
## 3:          97.00747             0           1          97.93340
## 4:          99.74761             1           0          98.40457
## 5:          99.57559             1           1          99.31678
## 6:          98.28889             1           1          99.82623
##     vomiting_day14 male
## 1:               1    0
## 2:               1    0
## 3:               0    1
## 4:               1    0
## 5:               1    0
## 6:               1    1
```

You are asked to analyze it. Patients' temperature and whether they are vomiting is recorded on day 0 of the experiment, then ZMapp is administered to patients in the treatment group on day 1. Vomiting and temperature is again recorded on day 14.

   a. Without using any covariates, answer this question with regression: What is the estimated effect of ZMapp (with standard error in parentheses) on whether someone was vomiting on day 14? What is the p-value associated with this estimate?

```
vomit_mod_1 <- d[, lm(vomiting_day14 ~ treat_zmapp)]
#use robust standard error
stargazer(
  vomit_mod_1,
  se = list(get_robust_se(vomit_mod_1)),
  type = 'text',
  report=('vc*sp'))
```

```
##
## ===============================================
##                      Dependent variable:
##                  ------------------------------
```

15

```
##                              vomiting_day14
## -----------------------------------------------
## treat_zmapp                     -0.238***
##                                  (0.091)
##                                 p = 0.010
##
## Constant                        0.847***
##                                  (0.048)
##                                 p = 0.000
##
## -----------------------------------------------
## Observations                      100
## R2                                0.073
## Adjusted R2                       0.063
## Residual Std. Error        0.421 (df = 98)
## F Statistic             7.705*** (df = 1; 98)
## ===============================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

The analysis estimates that we could expect the chance of vomiting on the 14th day to reduce by 23.8%. The p-value for this esitmate is 0.01

b. Add covariates for vomiting on day 0 and patient temperature on day 0 to the regression from part (a) and report the ATE (with standard error). Also report the p-value.

```
vomit_mod_2 <- d[, lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0)]

stargazer(
  vomit_mod_2,
  se = list(get_robust_se(vomit_mod_2)),
  type = 'text',
  report=('vc*sp'))
```

```
##
## ===============================================
##                       Dependent variable:
##                   ----------------------------
##                          vomiting_day14
## -----------------------------------------------
## treat_zmapp                     -0.166**
##                                  (0.082)
##                                 p = 0.044
##
## vomiting_day0                     0.065
##                                  (0.178)
##                                 p = 0.717
##
## temperature_day0                 0.206***
##                                  (0.078)
##                                 p = 0.009
##
## Constant                        -19.470**
##                                  (7.608)
##                                 p = 0.011
##
## -----------------------------------------------
```

```
## Observations                       100
## R2                               0.311
## Adjusted R2                       0.290
## Residual Std. Error        0.367 (df = 96)
## F Statistic            14.447*** (df = 3; 96)
## ===============================================
## Note:                 *p<0.1; **p<0.05; ***p<0.01
```

The figures are reported above.

c. Do you prefer the estimate of the ATE reported in part (a) or part (b)? Why? Report the results of the F-test that you used to form this opinion.

```r
anova(vomit_mod_1, vomit_mod_2, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: vomiting_day14 ~ treat_zmapp
## Model 2: vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     98 17.383
## 2     96 12.918  2    4.4653 16.592 6.472e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I would choose the estimate of the ATE from part (b) since the model in (b) does a better job at predicting the outcome even after penalizing the use of additional variables on the right hand side of the equation.

d. The regression from part (b) suggests that temperature is highly predictive of vomiting. Also include temperature on day 14 as a covariate in the regression from part (b) and report the ATE, the standard error, and the p-value.

```r
vomit_mod_3 <- d[, lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + temperature
```

```r
stargazer(
  vomit_mod_3,
  se = list(get_robust_se(vomit_mod_3)),
  type = 'text',
  report=('vc*sp'))
```

```
##
## ===============================================
##                      Dependent variable:
##                  ----------------------------
##                        vomiting_day14
## ---------------------------------------------
## treat_zmapp                 -0.120
##                             (0.086)
##                           p = 0.162
##
## vomiting_day0                0.046
##                             (0.173)
##                           p = 0.791
##
## temperature_day0            0.177**
##                             (0.077)
```

17

```
##                                 p = 0.022
##
## temperature_day14               0.060**
##                                 (0.026)
##                                 p = 0.020
##
## Constant                       -22.592***
##                                 (7.746)
##                                 p = 0.004
##
## -----------------------------------------------
## Observations                      100
## R2                               0.340
## Adjusted R2                      0.312
## Residual Std. Error      0.361 (df = 95)
## F Statistic           12.244*** (df = 4; 95)
## ===============================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

The figures are reported above.

e. Do you prefer the estimate of the ATE reported in part (b) or part (d)? Why? **Part (d) is the classic example of bad control. Since the temperature on day 14 could be affected by the treatment, including it on the right hand side of the equation potentially creates bias in our ATE estimation. Therefore, I would not choose the model in part (d).**

f. Now let's switch from the outcome of vomiting to the outcome of temperature, and use the same regression covariates as in part (b). Test the hypothesis that ZMapp is especially likely to reduce mens' temperatures, as compared to womens', and describe how you did so. What do the results suggest?

```
temp_mod_1 <- d[, lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + male + ma

stargazer(
  temp_mod_1,
  se = list(get_robust_se(temp_mod_1)),
  type = 'text')
```

```
##
## ===============================================
##                        Dependent variable:
##                    ----------------------------
##                         temperature_day14
## -----------------------------------------------
## treat_zmapp                  -0.231*
##                              (0.118)
##
## vomiting_day0                 0.041
##                              (0.195)
##
## temperature_day0            0.505***
##                              (0.105)
##
## male                        3.085***
##                              (0.122)
##
## treat_zmapp:male            -2.077***
```

```
##                               (0.198)
##
## Constant                      48.713***
##                               (10.194)
##
## ----------------------------------------------
## Observations                      100
## R2                               0.906
## Adjusted R2                      0.901
## Residual Std. Error       0.452 (df = 94)
## F Statistic          180.953*** (df = 5; 94)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

To test the hypothesis that ZMapp is especially likely to reduce men's temperatures, as compared to women's, we will need to look at the interaction between gender and treatment. It can be achieved by adding two more terms in our model, men and the interaction between men and treatment. From the coefficient for the interaction term, we can see that the drug has an additional effect to decrease the temprature by 2.077 degrees. Since the coefficient estimate is statistically significant, we can conclude that the treatment does reduce mens' temperatures more when compared to womens'.

g. Suspend reality for just a moment – suppose that you had the option of being a man or a woman who was a part of this study. Based on this data, which sex would you rather be? This time, you need to produce evidence (probably from your model estimates) to inform your determination. What does your determination depend on?

```
temp_mod_1 <- d[, lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + male + ma
vomit_mod_4 <- d[, lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + male + male
stargazer(
  temp_mod_1,
  vomit_mod_4,
  se = list(get_robust_se(temp_mod_1),
            get_robust_se(vomit_mod_4)),
  type = 'text')
```

```
##
## ===============================================================
##                            Dependent variable:
##                    --------------------------------
##                    temperature_day14 vomiting_day14
##                           (1)             (2)
## -------------------------------------------------------------
## treat_zmapp              -0.231*         -0.130
##                          (0.118)         (0.095)
##
## vomiting_day0             0.041           0.077
##                          (0.195)         (0.181)
##
## temperature_day0         0.505***        0.198**
##                          (0.105)         (0.080)
##
## male                     3.085***        0.016
##                          (0.122)         (0.091)
##
## treat_zmapp:male         -2.077***       -0.087
```

```
##                                        (0.198)           (0.169)
##
## Constant                              48.713***         -18.772**
##                                        (10.194)          (7.787)
##
## -----------------------------------------------------------------
## Observations                             100               100
## R2                                       0.906             0.314
## Adjusted R2                              0.901             0.277
## Residual Std. Error (df = 94)            0.452             0.370
## F Statistic (df = 5; 94)               180.953***         8.601***
## ===============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
```

Looking at the regression result above, we can see that gender is predictive of temperature but not vomiting on day 14. Therefore, I will focus the decision based on the temperature on day 14 (the result is still significant after we adjusted for the significance level by using Bonferroni Correction). Zoomming in on column 1, we can see that male is associated with higher temperature (3.085 degrees). Though there is also an additional treatment effect on male (-2.077 degrees), for a male receiving treatment, we could still expect the termperature to be higher than females' (3.085 - 2.077 > 0). Therefore, it seems that the overall severity of sympton is worse for male even when treatment effect is accounted for. So I would choose to be female in this study.

h. Suppose that you had not run the regression in part (f). Instead, you speak with a colleague to learn about heterogeneous treatment effects. This colleague has access to a non-anonymized version of the same dataset and reports that he had looked at heterogeneous effects of the ZMapp treatment by each of 10,000 different covariates to examine whether each predicted the effectiveness of ZMapp on each of 2,000 different indicators of health, for 20,000,000 different regressions in total. Across these 20,000,000 regressions your colleague ran, the treatment's interaction with gender on the outcome of temperature is the only heterogeneous treatment effect that he found to be statistically significant. He reasons that this shows the importance of gender for understanding the effectiveness of the drug, because nothing else seemed to indicate why it worked. Bolstering his confidence, after looking at the data, he also returned to his medical textbooks and built a theory about why ZMapp interacts with processes only present in men to cure. Another doctor, unfamiliar with the data, hears his theory and finds it plausible. How likely do you think it is ZMapp works especially well for curing Ebola in men, and why? (This question is conceptual can be answered without performing any computation.)

```
sig_level <- 0.05
num_regression <- 20000000
prob_at_least_one_sig <- 1-((1-sig_level)^20000000)
prob_at_least_one_sig
```

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

When looking at the calculation above, we see that the chance of identifying at least one significant interaction between sex and treatment when running so many regression is effectively one when the significance level is set at 0.05. Therefore, there is no proof whatsoever that ZMapp works especially well for curing Ebola in men.

i. Now, imagine that what described in part (g) did not happen, but that you had tested this heterogeneous treatment effect, and only this heterogeneous treatment effect, of your own accord. Would you be more or less inclined to believe that the heterogeneous treatment effect really exists? Why? **I would be more inclined to believe that the heterogeneous treatment effect does exist. Since only one effect was tested, the false positive rate is only 5%. That is there's only a 5% chance that the interaction discovered is due to chance alone. So, it's much more likely than not, that there is a real effect.**

j. Another colleague proposes that being of African descent causes one to be more likely to get Ebola. He asks you what ideal experiment would answer this question. What would you tell him? (*Hint: refer to Chapter 1 of Mostly Harmless Econometrics.*) **In order to answer this question in an experiment, we will have to randomly assign subjects to treatment so that treatment is independent of all of the omitted variables. This way any effect we see in the treatment group could be attributed to the treatment. However, the treatment in this question is African descent, and it is not an intervention that can be randomly assigned. There is no way we can find a group of subjects and randomly make some of them be of African descent. Therefore, this question cannot be answered by performing an experiment.**