

Problem Set 3

Experiments and Causality

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

Adding some libraries

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(sandwich)
```

```
library(stargazer)
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(multiwayvcov)
```

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.

1 Replicate Results

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

```
d <- read.csv("./data/broockman_green_anon_pooled_fb_users_only.csv")
d1 = fread("./data/broockman_green_anon_pooled_fb_users_only.csv")
```

- 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.

```
# take a look at the data
head(d1)
```

```
##      studyno treat_ad      cluster name_recall
## 1:         2         0 Study 2, Cluster Number 1      0
## 2:         2         0 Study 2, Cluster Number 2      1
## 3:         2         0 Study 2, Cluster Number 3      0
## 4:         2         0 Study 2, Cluster Number 4      1
## 5:         2         1 Study 2, Cluster Number 7      1
## 6:         2         1 Study 2, Cluster Number 7      0
##      positive_impression
## 1:                      0
## 2:                      0
## 3:                      0
## 4:                      0
## 5:                      1
## 6:                      0
```

```
# summarize the data
(summary(d1))
```

```
##      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
```

```
# create a linear regression model for study 1
modell1 = lm(name_recall ~ treat_ad, data = d1[studyno == 1])
```

```
stargazer(model1)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:15

Table 1:

<i>Dependent variable:</i>	
	name_recall
treat_ad	-0.010 (0.021)
Constant	0.182*** (0.016)
Observations	1,364
R ²	0.0002
Adjusted R ²	-0.001
Residual Std. Error	0.382 (df = 1362)
F Statistic	0.217 (df = 1; 1362)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
model1_conf = confint(model1, level = 0.95)
cat("The 95% confidence interval is:", model1_conf["treat_ad",])
```

```
## The 95% confidence interval is: -0.05101765 0.03142188
```

- b. What are the clusters in Broockman and Green's study? Why might taking clustering into account increase the standard errors?

Clusters in the Broockman and Green study consist of individuals with unique combinations of age, gender, and location. One example could be twenty-four year old males in San Francisco, California. Standard errors are inversely proportional to the variance in the input variable. Since we are reducing the variance in the input variable by clustering, we will increase our standard errors.

- c. Now repeat part (a), but taking clustering into account. That is, compute a confidence interval for the effect of the ad on candidate name recognition in Study 1, but now correctly accounting for the clustered nature of the treatment assignment. 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: `./code/week5clusterAndRobust.Rmd`.

```
model1_cluster_vcv = cluster.vcov(model1, ~ cluster)
model1_cluster_se = sqrt(diag(model1_cluster_vcv))
model1_res = c(model1$coefficients["treat_ad"] - 1.96*model1_cluster_se, model1$coefficients["treat_ad"] + 1.96*model1_cluster_se)
cat("The confidence interval is between:", model1_res[2], "and", model1_res[4])
```

```
## The confidence interval is between: -0.05635499 and 0.03675922
```

- d. Repeat part (c), but now for Study 2 only.

```
model2 = lm(name_recall ~ treat_ad, data = d1[studyno == 2])
```

```
stargazer(model2)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Tue, Feb 26, 2019 - 00:22:16

Table 2:

	<i>Dependent variable:</i>
	name_recall
treat_ad	-0.003 (0.031)
Constant	0.606*** (0.015)
Observations	1,337
R ²	0.00001
Adjusted R ²	-0.001
Residual Std. Error	0.489 (df = 1335)
F Statistic	0.008 (df = 1; 1335)

Note: *p<0.1; **p<0.05; ***p<0.01

```
model2_cluster_vcv = cluster.vcov(model2, ~ cluster)
model2_cluster_se = sqrt(diag(model2_cluster_vcv))
model2_res = c(model2$coefficients["treat_ad"] - 1.96*model2_cluster_se, model2$coefficients["treat_ad"] + 1.96*model2_cluster_se)
cat("The confidence interval is between:", model2_res[2], "and", model2_res[4])
```

The confidence interval is between: -0.0723899 and 0.0667832

- e. Repeat part (c), but using 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 and run one omnibus regression. What is the treatment effect estimate and associated p-value?

```
model1e = lm(name_recall ~ treat_ad, data = d1)
```

```
stargazer(model1e)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Tue, Feb 26, 2019 - 00:22:16

```
model1e_cluster_vcv = cluster.vcov(model1e, ~ cluster)
model1e_cluster_se = sqrt(diag(model1e_cluster_vcv))
model1e_res = c(model1e$coefficients["treat_ad"] - 1.96*model1e_cluster_se, model1e$coefficients["treat_ad"] + 1.96*model1e_cluster_se)
cat("The ATE of the treatment effect is:", model1e$coefficients[2], "\n")
```

The ATE of the treatment effect is: -0.1550732

```
cat("The associated p-value is:", summary(model1e)$coefficients[2,4])
```

The associated p-value is: 2.160639e-16

- f. Now, repeat part (e) but include a dummy variable (a 0/1 binary variable) for whether the data are from Study 1 or Study 2. What is the treatment effect estimate and associated p-value?

```
d1[, study_var := ifelse(studyno == 1, 0, 1)]
```

```
model1f = lm(name_recall ~ treat_ad + study_var, data = d1)
```

Table 3:

	<i>Dependent variable:</i>
	name_recall
treat_ad	-0.155*** (0.019)
Constant	0.454*** (0.012)
Observations	2,701
R ²	0.025
Adjusted R ²	0.024
Residual Std. Error	0.482 (df = 2699)
F Statistic	68.313*** (df = 1; 2699)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

```
stargazer(model1f)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Tue, Feb 26, 2019 - 00:22:16

Table 4:

	<i>Dependent variable:</i>
	name_recall
treat_ad	-0.007 (0.018)
study_var	0.426*** (0.018)
Constant	0.181*** (0.016)
Observations	2,701
R ²	0.193
Adjusted R ²	0.193
Residual Std. Error	0.438 (df = 2698)
F Statistic	322.848*** (df = 2; 2698)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

```
model1f_cluster_vcv = cluster.vcov(model1f, ~ cluster)
model1f_cluster_se = sqrt(diag(model1f_cluster_vcv))
model1f_res = c(model1f$coefficients["treat_ad"] - 1.96*model1f_cluster_se, model1f$coefficients["treat_
cat("The ATE of the treatment effect is:", model1f$coefficients[2], "\n")
```

```
## The ATE of the treatment effect is: -0.006775249
```

```
cat("The associated p-value is:", summary(model1f)$coefficients[2,4])
```

```
## The associated p-value is: 0.7093688
```

- g. Why did the results from parts (e) and (f) differ? Which result is biased, and why? (Hint: see pages 75-76 of Gerber and Green, with more detailed discussion optionally available on pages 116-121.)

The results differed because there are inherent differences in study 1 vs study 2. The result from part e is biased because it does not account for these differences in the regression. Firstly, the probability of receiving treatment being different between the studies. Additionally in part f, we include a covariate which allows us to block on the study type and gives a less biased result.

- h. Skim this Facebook case study and consider two claims they make reprinted below. Why might their results differ from Broockman and Green's? Please be specific and provide examples.
- "There was a 19 percent difference in the way people voted in areas where Facebook Ads ran versus areas where the ads did not run."
 - **The Facebook ads ran in heavily populated areas and comparisons were made to lesser populated areas. It almost feels like there was a selection bias in play. Because of this, the results of the comparison cannot be trusted.**
 - "In the areas where the ads ran, people with the most online ad exposure were 17 percent more likely to vote against the proposition than those with the least."
 - **Candidates for the study were chosen using specified criteria. Because we do not know how the conductors of the experiment accounted for these (i.e. was blocking/clustering used?) we have no way of knowing how reliable the results are. The experimenters also selected subjects that expressed interest in the political sector. This inherently introduces a selection bias.**

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.

```
ate_2 = 0.187
se_2 = 0.032
conf_int_2 = c(ate_2 - 1.96*se_2, ate_2 + 1.96*se_2)
cat("The estimated ATE is:", ate_2, "\n")
```

```
## The estimated ATE is: 0.187
```

```
cat("The 95% confidence interval is:", conf_int_2)
```

```
## The 95% confidence interval is: 0.12428 0.24972
```

- 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.

```
ate_2 = -0.024
se_2 = 0.039
conf_int_2 = c(ate_2 - 1.96*se_2, ate_2 + 1.96*se_2)
cat("The estimated ATE is:", ate_2, "\n")
```

```
## The estimated ATE is: -0.024
```

```
cat("The 95% confidence interval is:", conf_int_2)
```

```
## The 95% confidence interval is: -0.10044 0.05244
```

- 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 - 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 of the outcome measures in Table 4A show statistically significant effects (at the 5% level) of sending a text message.

- 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.

```
avg_weight_per_week = 0.281
weight_delta = 2
est_rec = avg_weight_per_week*weight_delta
cat("We expect household A to have", est_rec, "kg more recycling per week than household B")
```

We expect household A to have 0.562 kg more recycling per week than household B

- 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.

We would not expect the ATE to change assuming that the experiment has been randomized properly. The standard error would increase as we are reducing the variance in the input variable by removing the baseline.

- g. In column 1 of Table 4A, would you say the variable “has cell phone” is a bad control? Explain your reasoning.

I think it is a good control primarily because it has very little impact on the experiment itself. Presence or lack of a cell-phone is not likely to effect how/if you recycle. It also gives us the ability to understand the *Any SMS Message* coefficient Since the variable would not hold any meaning for those without cellphones and allows us to filter those users properly.

- 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.

Thinking logically, recieving an sms message can be thought of as being independent of whether or not someone has a cellphone or not. Hence, these two variables would not be correlated (assuming random assignment) and there should be no impact on “Any SMS message”.

3 Multifactor Experiments

Staying with the same experiment, now let's 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.)

This can be considered a 3x3 study. The breakdown is as follows: [bin w/ sticker, bin w/o sticker, no bin (control)] by [personal sms message, generic sms message, no sms message (control)]

- 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 measurements were obtained during the first two weeks of the data collection period. They are considered the baseline because no bins were distributed nor were any text messages sent. Hence, the baseline can be thought of a measurement of the outcome variables before any treatment is applied. Doing this allows us to obtain a proper frame of reference for the outcome results post-treatment.

- c. In column (1) of Table 4B, interpret the magnitude of the coefficient on “bin without sticker.” What does it mean?

The *bin without sticker* coefficient under the *percentage of visits turned in bag* represents the treatment effect the former had on the latter. With that being said, the values of *0.035* indicates that we saw a statistically significant increase in *percentage of visits turned in bag* when the treatment of *bin without sticker* was applied.

- 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?

The *with sticker*(0.055) seems to have a stronger treatment effect than the *no sticker*(0.035). The delta in the magnitude is 0.02.

- e. Is this difference you just described statistically significant? Explain which piece of information in the table allows you to answer this question.

Given that both treatments have the same standard error, we can see that 0.02 (delta) > 0.015 (standard error) and hence is not 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.”

The model is considered saturated in the sense that it contains all the covariates as well as their interaction terms.

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).

```
d <- read.dta("./data/karlan_data_subset_for_class.dta")
d4 = data.table(d)
head(d4)
```

```
##      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
```

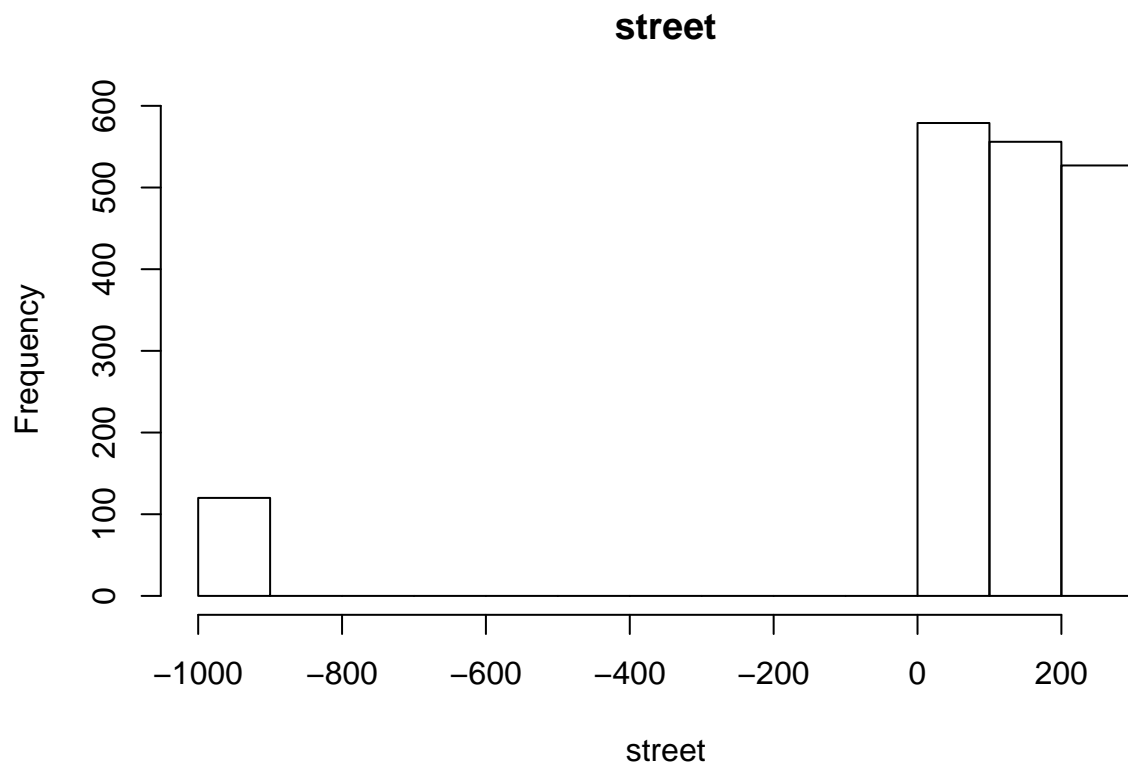
Do some quick exploratory data analysis with this data. There are some values in this data that seem

```
summary(d4)
```

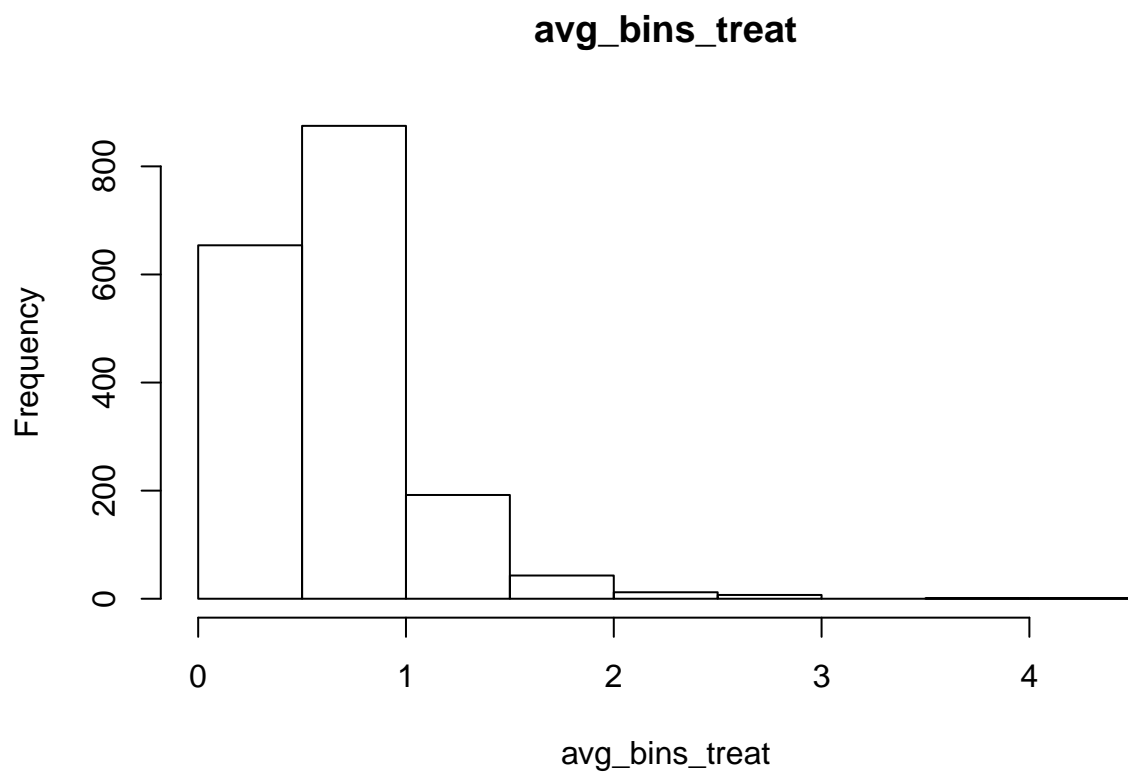
```
##      street      havecell      avg_bins_treat      base_avg_bins_treat
## Min.   :-999.00   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:  69.00   1st Qu.:0.0000   1st Qu.:0.4167   1st Qu.:0.3750
## Median : 131.50   Median :1.0000   Median :0.6250   Median :0.6250
## Mean   :  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
##
```

The minimum value for *street* is -999 which seems incorrect. It might be that these values were not collected during the data collection phase. Let's take a deeper look at the distribution of non-binary data.

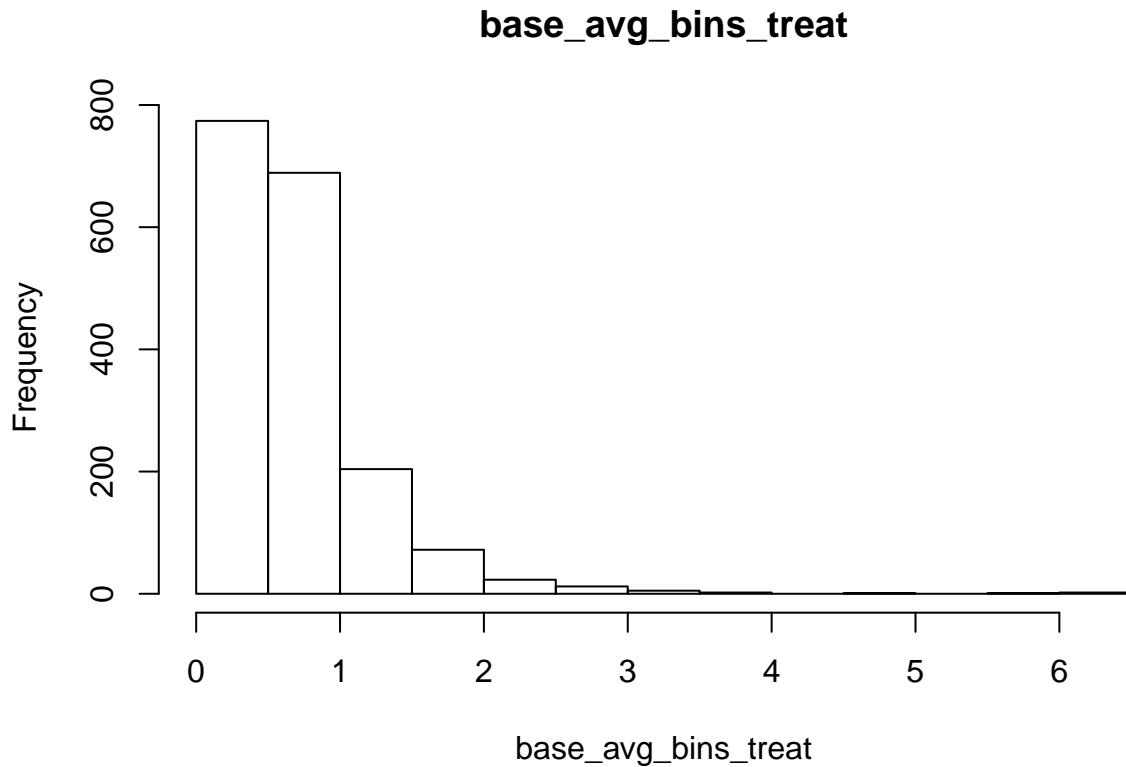
```
hist(d4[, street], main = "street", xlab = "street")
```



```
hist(d4[, avg_bins_treat], main = "avg_bins_treat", xlab = "avg_bins_treat")
```



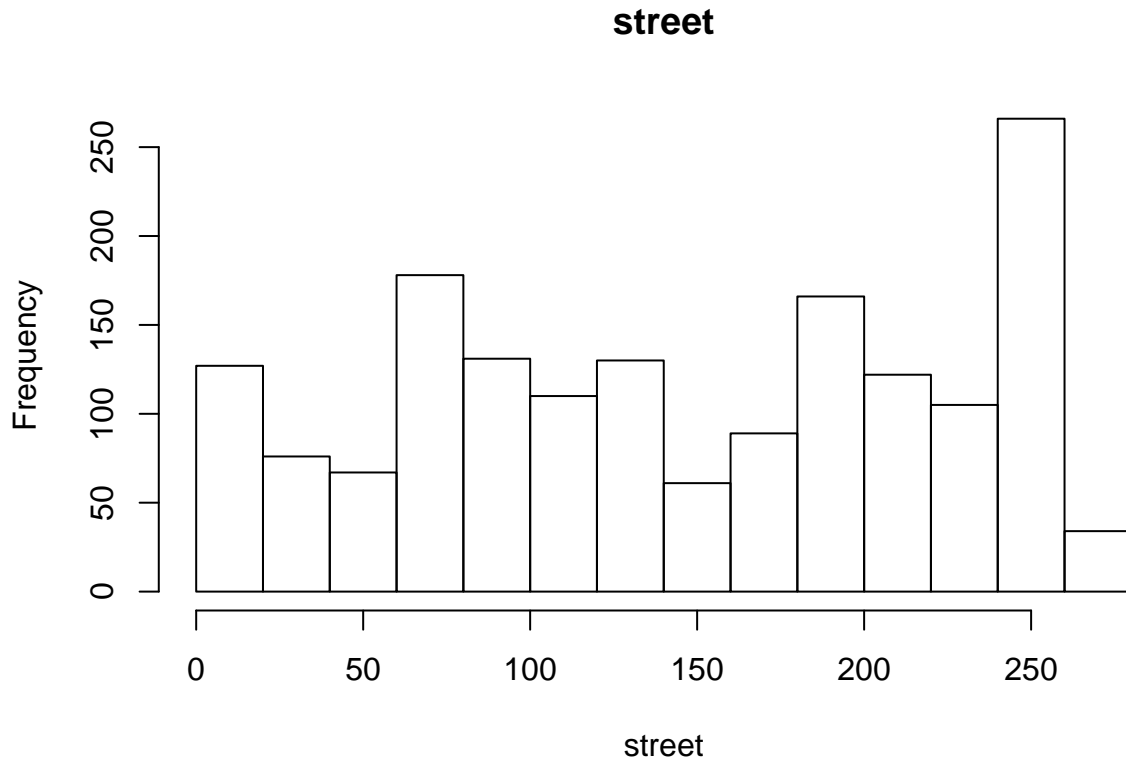
```
hist(d4[, base_avg_bins_treat], main = "base_avg_bins_treat", xlab = "base_avg_bins_treat")
```



The distributions for `avg_bins_treat` and `base_avg_bins_treat` are skewed but the data looks correct. We can clearly see that we have a about a hundred values for *street* that are -999. Knowing this we could just account for them in all our calculations. However, it would be better practice to go ahead and convert these to NA. That way if we don't account for them properly we will get something along the lines of an incompatible type error.

```
# replace -999 with NA
d4[street == -999] = NA
```

```
# sanity check
hist(d4[, street], main = "street", xlab = "street")
```



Let's take a look at the binary variables. Since we have no way to whether the one or zero is valid a good way to validate the data is to see the total number of ones and zeros and corss-reference them with the total number of rows in our data table.

```
count = d4[, .N]
havecell_count = d4[havecell == 0 || havecell == 1, .N]
if (count != havecell_count)
  stop("The count for 'havecell' is not correct! Please double-check the data!")
bin_count = d4[bin == 0 || bin == 1, .N]
if (count != bin_count)
  stop("The count for 'bin' is not correct! Please double-check the data!")
bins_count = d4[bin_s == 0 || bin_s == 1, .N]
if (count != bins_count)
  stop("The count for 'bin_s' is not correct! Please double-check the data!")
bing_count = d4[bin_g == 0 || bin_g == 1, .N]
if (count != bing_count)
  stop("The count for 'bin_g' is not correct! Please double-check the data!")
sms_count = d4[sms == 0 || sms == 1, .N]
if (count != sms_count)
  stop("The count for 'sms' is not correct! Please double-check the data!")
smssp_count = d4[sms_p == 0 || sms_p == 1, .N]
if (count != smssp_count)
  stop("The count for 'sms_p' is not correct! Please double-check the data!")
smsg_count = d4[sms_g == 0 || sms_g == 1, .N]
if (count != smsg_count)
  stop("The count for 'sms_g' is not correct! Please double-check the data!")

cat("The binary variables match up with the total number of data points!")
```

```
## The binary variables match up with the total number of data points!
```

- 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.

```
model4_a = lm(avg_bins_treat ~ bin, data = d4)
```

```
stargazer(model4_a)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:17

Table 5:

	<i>Dependent variable:</i>
	avg_bins_treat
bin	0.137*** (0.021)
Constant	0.635*** (0.012)
Observations	1,665
R ²	0.025
Adjusted R ²	0.025
Residual Std. Error	0.406 (df = 1663)
F Statistic	42.980*** (df = 1; 1663)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

```
model4_a_confint = confint(model4_a, level = 0.95)
cat("The confidence interval is:", model4_a_confint["bin",])
```

```
## The confidence interval is: 0.09610603 0.1781613
```

- 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.

```
model4_b = lm(avg_bins_treat ~ bin + base_avg_bins_treat, data = d4)
```

```
stargazer(model4_b)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:17

```
model4_b_confint = confint(model4_b, level = 0.95)
cat("The confidence interval is:", model4_b_confint["bin",])
```

```
## The confidence interval is: 0.09387119 0.1613531
```

The first thing to note here is that the width of the confidence interval shrinks. This is because by adding the baseline the variance of the outcome is reduced/explained. Reduction in variance in the outcome reduces our standard error resulting in more accurate confidence intervals.

- 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.

Table 6:

	<i>Dependent variable:</i>
	avg_bins_treat
bin	0.128*** (0.017)
base_avg_bins_treat	0.389*** (0.014)
Constant	0.352*** (0.014)
Observations	1,665
R ²	0.341
Adjusted R ²	0.341
Residual Std. Error	0.334 (df = 1662)
F Statistic	430.682*** (df = 2; 1662)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
model4_c = lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street), data = d4)
```

```
stargazer(model4_c)
```

```
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:17
```

```
model4_c_confint = confint(model4_c, level = 0.95)
cat("The confidence interval is:", model4_c_confint["bin",])
```

```
## The confidence interval is: 0.08175545 0.1507503
```

- 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.

We do not see a significant difference between the confidence intervals between the two previous parts. We would expect that blocking would help improve the precision of our estimates if we expected the covariates to vary by street. This leads us to believe that the coefficients of the models do not vary from block to block.

- 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.

```
d4[, 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.

```
model4_f = lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street) + no_cell_phone, data = d4)
```

```
stargazer(model4_f)
```

Table 7:

	<i>Dependent variable:</i>
	avg_bins_treat
bin	0.116*** (0.018)
base_avg_bins_treat	0.367*** (0.015)
factor(street)3	0.051 (0.141)
factor(street)4	−0.041 (0.145)
factor(street)5	0.001 (0.145)
factor(street)6	0.044 (0.150)
factor(street)7	0.039 (0.174)
factor(street)8	0.041 (0.130)
factor(street)9	0.008 (0.135)
factor(street)10	0.068 (0.122)
factor(street)11	0.027 (0.135)
factor(street)15	0.074 (0.189)
factor(street)17	0.135 (0.164)
factor(street)20	−0.138 (0.150)
factor(street)21	0.018 (0.164)
factor(street)22	−0.030 (0.116)
factor(street)23	−0.064 (0.118)
factor(street)26	0.070 (0.211)
factor(street)32	0.250

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:17

```
model4_f_confint = confint(model4_f, level = 0.95)  
cat("The confidence interval is:", model4_f_confint["bin",])
```

The confidence interval is: 0.082673 0.1516659

We do not see a significant impact on the confidence interval of the treatment variable when using “no cell phone” vs “has cell phone”. This makes sense because we do not expect that possession, or lack thereof, of a cellphone effects if a bin is distributed resulting in more recycling. This variable is useful however, if we want to analyze the effect of sending an sms message as we will do later.

- 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.

```
model4_g = lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street) + no_cell_phone + sms, data =  
stargazer(model4_g)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:18

```
model4_g_confint = confint(model4_g, level = 0.95)  
cat("The confidence interval is:", model4_g_confint["bin",])
```

The confidence interval is: 0.0824637 0.1514719

- 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.

```
model4_h = lm(avg_bins_treat ~ bin_g + bin_s + sms_g + sms_p + base_avg_bins_treat + no_cell_phone + fa  
stargazer(model4_h)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:18

```
model4_h_confint = confint(model4_h, level = 0.95)  
cat("The confidence interval is:", model4_h_confint["bin_g",])
```

The confidence interval is: 0.06116499 0.1503123

The confidence interval in part h is a little tighter than in part g. Instead on the bin variable we now have three variables: bin_g and bin_s. This allows us to explain more of the variance resulting in a tighter confidence interval.

Table 8:

	<i>Dependent variable:</i>	
	avg_bins_treat	
bin	0.117***	(0.018)
base_avg_bins_treat	0.367***	(0.015)
factor(street)3	0.047	(0.141)
factor(street)4	−0.029	(0.145)
factor(street)5	−0.001	(0.145)
factor(street)6	0.043	(0.150)
factor(street)7	0.027	(0.174)
factor(street)8	0.044	(0.130)
factor(street)9	0.011	(0.135)
factor(street)10	0.070	(0.122)
factor(street)11	0.026	(0.135)
factor(street)15	0.073	(0.189)
factor(street)17	0.138	(0.164)
factor(street)20	−0.128	(0.150)
factor(street)21	0.042	(0.164)
factor(street)22	−0.018	(0.116)
factor(street)23	−0.054	(0.118)
factor(street)26	0.087	(0.210)
factor(street)32	0.267	

Table 9:

	<i>Dependent variable:</i>
	avg_bins_treat
bin	0.117*** (0.018)
base_avg_bins_treat	0.367*** (0.015)
factor(street)3	0.046 (0.141)
factor(street)4	−0.031 (0.145)
factor(street)5	−0.002 (0.145)
factor(street)6	0.041 (0.150)
factor(street)7	0.026 (0.174)
factor(street)8	0.044 (0.130)
factor(street)9	0.011 (0.135)
factor(street)10	0.067 (0.122)
factor(street)11	0.024 (0.135)
factor(street)15	0.075 (0.189)
factor(street)17	0.137 (0.164)
factor(street)20	−0.130 (0.150)
factor(street)21	0.040 (0.164)
factor(street)22	−0.019 (0.116)
factor(street)23	−0.055 (0.118)
factor(street)26	0.083 (0.211)
factor(street)32	0.264

Table 10:

	<i>Dependent variable:</i>
	avg_bins_treat
bin_g	0.106*** (0.023)
bin_s	0.129*** (0.023)
sms_g	0.030 (0.026)
sms_p	0.007 (0.026)
base_avg_bins_treat	0.367*** (0.015)
no_cell_phone	-0.033 (0.021)
factor(street)3	0.046 (0.141)
factor(street)4	-0.034 (0.145)
factor(street)5	-0.003 (0.145)
factor(street)6	0.043 (0.150)
factor(street)7	0.025 (0.174)
factor(street)8	0.043 (0.130)
factor(street)9	0.009 (0.135)
factor(street)10	0.068 (0.122)
factor(street)11	0.021 (0.135)
factor(street)15	0.068 (0.189)
factor(street)17	0.132 (0.164)
factor(street)20	-0.130 (0.150)
factor(street)21	0.031

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 <- read.csv("./data/ebola_rct2.csv")
d5 = fread("./data/ebola_rct2.csv")
head(d5)

##      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?

```
model5_a = lm(vomiting_day14 ~ treat_zmapp, data = d5)
```

```
stargazer(model5_a)
```

```
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:18
```

```
cat("The ATE of treat_zmapp is:", model5_a$coefficients[2], "(", summary(model5_a)$coefficients[2,2], "
```

```
## The ATE of treat_zmapp is: -0.2377015 ( 0.08563161 )
```

```
cat("The associated p-value is:", summary(model5_a)$coefficients[2,4])
```

```
## The associated p-value is: 0.006595412
```

- 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.

```
model5_b = lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0, data = d5)
```

```
stargazer(model5_b)
```

```
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:18
```

Table 11:

	<i>Dependent variable:</i>
	vomiting__day14
treat_zmapp	−0.238*** (0.086)
Constant	0.847*** (0.055)
Observations	100
R ²	0.073
Adjusted R ²	0.063
Residual Std. Error	0.421 (df = 98)
F Statistic	7.705*** (df = 1; 98)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 12:

	<i>Dependent variable:</i>
	vomiting__day14
treat_zmapp	−0.166** (0.076)
vomiting__day0	0.065 (0.146)
temperature__day0	0.206*** (0.076)
Constant	−19.470** (7.441)
Observations	100
R ²	0.311
Adjusted R ²	0.290
Residual Std. Error	0.367 (df = 96)
F Statistic	14.447*** (df = 3; 96)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

```
cat("The ATE of treat_zmapp is:", model5_b$coefficients[2], "(", summary(model5_a)$coefficients[2,2], ")")
## The ATE of treat_zmapp is: -0.1655367 ( 0.08563161 )
cat("The associated p-value is:", summary(model5_b)$coefficients[2,4])
## The associated p-value is: 0.03112852
```

c. Do you prefer the estimate of the ATE reported in part (a) or part (b)? Why?

We would prefer the ATE from part b. Firstly we see that the addition of temperature_day0 is statistically significant, hence, the inclusion of it is important. We see that vomiting_day0 is not statistically significant as per the regression so would likely see the same results regardless of its inclusion.

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.

```
model5_d = lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + temperature_day14, data = data)
stargazer(model5_d)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Tue, Feb 26, 2019 - 00:22:18

Table 13:	
	<i>Dependent variable:</i>
	vomiting_day14
treat_zmapp	-0.120 (0.078)
vomiting_day0	0.046 (0.144)
temperature_day0	0.177** (0.076)
temperature_day14	0.060** (0.029)
Constant	-22.592*** (7.477)
Observations	100
R ²	0.340
Adjusted R ²	0.312
Residual Std. Error	0.361 (df = 95)
F Statistic	12.244*** (df = 4; 95)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

```
cat("The ATE of treat_zmapp is:", model5_d$coefficients[2], "(", summary(model5_d)$coefficients[2,2], ")")
## The ATE of treat_zmapp is: -0.1201006 ( 0.07767979 )
```

```
cat("The associated p-value is:", summary(model5_d)$coefficients[2,4])
```

```
## The associated p-value is: 0.1254056
```

e. Do you prefer the estimate of the ATE reported in part (b) or part (d)? Why?

We would still prefer the ATE reported in part b. This is because temperature_day14 is measured post-treatment and, hence, would not be a good control.

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 men's temperatures, as compared to women's, and describe how you did so. What do the results suggest?

```
model5_f = lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + male + male*treat_zmapp)
```

```
stargazer(model5_f)
```

```
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Feb 26, 2019 - 00:22:19
```

Table 14:	
	<i>Dependent variable:</i>
	temperature_day14
treat_zmapp	-0.231* (0.119)
vomiting_day0	0.041 (0.182)
temperature_day0	0.505*** (0.095)
male	3.085*** (0.126)
treat_zmapp:male	-2.077*** (0.192)
Constant	48.713*** (9.266)
Observations	100
R ²	0.906
Adjusted R ²	0.901
Residual Std. Error	0.452 (df = 94)
F Statistic	180.953*** (df = 5; 94)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

```
cat("The ATE of treat_zmapp is:", model5_f$coefficients[2], "(", summary(model5_f)$coefficients[2,2], ")")
```

```
## The ATE of treat_zmapp is: -0.2308655 ( 0.11871 )
```

```
cat("The associated p-value is:", summary(model5_f)$coefficients[2,4])
```


The associated p-value is: 0.05478966

We start by including the “male” dummy variable in our regression. We also add an interaction term between “male” and “treat_zapp”. This way we can measure the effect of being male and it effecting temperature as well as being a male who received treatment and it effecting temperature. Analyzing the table, a male who has received treatment sees an overall delta of -2 units in temperature on the 14th day. While being just male sees an overall delta of +3 units in temperature on the 14th day.

- g. Suppose that you had not run the regression in part (f). Instead, you speak with a colleague to learn about heterogenous treatment effects. This colleague has access to a non-anonymized version of the same dataset and reports that he had looked at heterogenous 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 heterogenous 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.)

There is not enough information within the paragraph for me to concretely believe that ZMapp works especially well for curing ebola in men. Within 20,000,000 regressions you would expect different variables to show up as being statistically significant. How many times did our colleague observe the treatment’s interaction term come out as statistically significant?

- h. 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 really exists if I was able to see it as being statistically significant while running a systematic regression centered around the heterogeneous effect.

- i. 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.*)

I would tell him he is crazy and needs to go read Field Experiments because this kind of experiment is impossible in terms of practicality

In this experiment we would get a randomized sample of the population and divide them into treatment and control. We would then expose the treatment group to a curable strain of the ebola virus. The more and ethical lines that are being crossed here cannot be understated. This cannot be an observational study because that will not for certain prove to us the existence of a causal claim. After the exposure we would monitor the subjects and block on being of African descent versus not. After this we can analyze the results to know for certain whether there is a causal relationship. Let’s not forget to administer the cure...