Problem Set 3

Experiments and Causality

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

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

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

```
##  R package version 5.2.1. https://CRAN.R-project.org/package=stargazer
```

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.

```
robust_model = function(model, cluster = NULL) {
    model$vcovHC = vcovHC(model)
    if (!is.null(cluster))
        model$cluster.vcov = cluster.vcov(model, cluster)
    model
}
```

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

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.

```
d1 = d[grepl('Study 1', cluster), ]
m1 = robust_model(d1[, lm(name_recall ~ treat_ad)], d1$cluster)
coefci(m1)
```

```
##                   2.5 %     97.5 %
## (Intercept)  0.15080247 0.21413492
## treat_ad    -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?

The study had clusters of people grouped by their demographics, e.g. age, gender, and location. People within the same cluster may respond very similarly, but the intercluster variance may be high. Since there is no way to randomize within clusters, they act like a single data point each, increasing the 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`.

```
coefci(m1, vcov. = m1$cluster.vcov)
```

```
##                   2.5 %     97.5 %
## (Intercept)  0.14619376 0.21874363
## treat_ad    -0.05639555 0.03679977
```

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

```
d2 = d[grepl('Study 2', cluster), ]
m2 = robust_model(d2[, lm(name_recall ~ treat_ad)], d2$cluster)
coefci(m2, vcov. = m2$cluster.vcov)
```

```
##                   2.5 %     97.5 %
## (Intercept)  0.57010643 0.64147042
## treat_ad    -0.07245176 0.06684507
```

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?

```
m3 = robust_model(d[, lm(name_recall ~ treat_ad)], d$cluster)
(co = coeftest(m3, m3$cluster.vcov))
```

```
##
## 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
```

```
# treatment effect
co[2,1]
```

```
## [1] -0.1550732
```

```
# p value
co[2,4]
```

```
## [1] 7.343954e-09
```

```
# confidence interval
coefci(m3, vcov. = m3$cluster.vcov)
```

```
##                   2.5 %     97.5 %
## (Intercept)  0.4177709  0.4906211
## treat_ad    -0.2074875 -0.1026589
```

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?

```
m4 = robust_model(d[, lm(name_recall ~ treat_ad + studyno)], d$cluster)
(co = coeftest(m4, m4$cluster.vcov))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -0.2454140  0.0340571 -7.2059 7.447e-13 ***
## treat_ad    -0.0067752  0.0204154 -0.3319      0.74
## studyno      0.4260988  0.0206970 20.5875 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# treatment effect
co[2,1]
```

```
## [1] -0.006775249
```

```
# p value
co[2,4]
```

```
## [1] 0.7400138
```

```
# confidence interval
coefci(m4, vcov. = m4$cluster.vcov)
```

```
##                   2.5 %     97.5 %
## (Intercept) -0.3121948 -0.1786333
## treat_ad    -0.0468067  0.0332562
## studyno      0.3855153  0.4666823
```

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 researchers had two different studies. Several factors differed between them, including the candidates' party affliation and their viabilities to win. Part f shows these factors are much better predictors of whether people remembered their names than whether or not they saw the ads. Part e is biased because it did not include the study number.

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."
- "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."

Facebook applied their treatment to areas that were most likely to be affected by the treatment. There is not mention in the article of any randomization or control group. It is not clear if this was an actual experiment.

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 average treatment effect is 0.187 kg with a confidence interval of (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 average treatment effect is -0.024 kg with a confidence interval of (-0.102, 0.054).

c. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of providing a recycling bin?

All of them show statistically significant effects except for average percentage of contamination per week.

d. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of sending text messages?

None of them show statistically significant effects.

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.

Household A would turn in 0.562 kg 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.

Not including the baseline variable should not effect the ATE since the treatment assignment was randomized. However, the standard error would increase since the variability explained by the baseline variable would have to be absorbed into other correlated variables.

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

The treatment is unlikely to have any effect on whether people obtain cell phones, so in that sense, it is not bad conctrol. However, not having a cell phone precludes a person from receiving the SMS message. In this sense, it is a bad control because having a cell phone is heavily correlated with being in teh treatment group.

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.

The coefficient would go up because getting an SMS message is highly correlated with having a cell phone. If the cell phone indicator is removed, its effect would be absorbed into the SMS variable.

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.)

The experimental design is 3x3. There are two treatments and a control for bins; two treatments and a control for text messages. It can be argued that there is an additional 3x1 study on people who do not have cell phones.

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 category is people who received no bin or text messages.

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

Compared to the baseline, people who receive a bin without a sticker turned in recycling on 3.5% more of visits.

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 bin with a sticker has a larger effect, by 2.0%.

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

The difference is not statistically significant because the standard error on both effects is 1.5%, which will clearly cause the two 95% confidence intervals to overlap. In addition the author's performed an F-test that showed a p-value of 0.31 for difference between the two models.

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 models consider every possible interaction term. Table 4C has every combination of bins and text message treatments.

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")
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
```

```
d = data.table(d)

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

The street variable has a -999 value on 120 rows. There are also 3 NAs in street and 1 in havecell. Judging from the number of rows in the original study's regression tables, the author left in the -999 street but removed the NAs.

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.

```
m = d[, lm(avg_bins_treat ~ bin)]
summary(m)
```

```
##
## Call:
## lm(formula = avg_bins_treat ~ bin)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7707 -0.2603 -0.0520  0.1876  3.5313
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.63535    0.01179  53.874  < 2e-16 ***
## bin          0.13538    0.02029   6.672 3.36e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4055 on 1783 degrees of freedom
## Multiple R-squared:  0.02436,    Adjusted R-squared:  0.02381
## F-statistic: 44.52 on 1 and 1783 DF,  p-value: 3.356e-11
```

```
coefci(m)
```

```
##                   2.5 %    97.5 %
## (Intercept) 0.61221964 0.6584797
```

```
## bin         0.09558421 0.1751758
```

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.

```
m = d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat)]
summary(m)
```

```
##
## Call:
## lm(formula = avg_bins_treat ~ bin + base_avg_bins_treat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.01396 -0.21275 -0.02647  0.16665  2.13549
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.34960    0.01373  25.460  < 2e-16 ***
## bin                  0.12469    0.01667   7.481 1.15e-13 ***
## base_avg_bins_treat  0.39296    0.01339  29.356  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.333 on 1782 degrees of freedom
## Multiple R-squared:  0.3424, Adjusted R-squared:  0.3416
## F-statistic: 463.9 on 2 and 1782 DF,  p-value: < 2.2e-16
```

```
coefci(m)
```

```
##                          2.5 %    97.5 %
## (Intercept)         0.32267158 0.3765335
## bin                 0.09200378 0.1573822
## base_avg_bins_treat 0.36671039 0.4192189
```

This confidence interval differs from the previous one (it's narrower) because adding the baseline number of bins increased the precision at which we estimate the ATE.

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.

Regression with -999 street removed.

```
m = d[street != -999, lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street))]
# summary(m) suppressed because there's 180 different streets
coef(summary(m))[1:3,]
```

```
##                      Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)         0.2760730 0.09789442  2.820110  4.864696e-03
## bin                 0.1162529 0.01758668  6.610282  5.338231e-11
## base_avg_bins_treat 0.3669966 0.01476526 24.855413 2.878717e-114
```

```
coefci(m, vcov. = cluster.vcov(m, d[street != -999, street]))[1:3,]
```

```
##                          2.5 %    97.5 %
## (Intercept)         0.24134695 0.3107990
## bin                 0.07763493 0.1548709
## base_avg_bins_treat 0.30445608 0.4295372
```

Regerssion with -999 street included.

```
m = d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street))]
# summary(m) suppressed because there's 180 different streets
coef(summary(m))[1:3,]
```

```
##                       Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)          0.3677440 0.03161561 11.631723  4.442673e-30
## bin                  0.1138868 0.01705784  6.676508  3.364562e-11
## base_avg_bins_treat 0.3737068 0.01432809 26.082099 2.936935e-125
```

```
coefci(m, vcov. = cluster.vcov(m, d$street))[1:3,]
```

```
##                          2.5 %    97.5 %
## (Intercept)          0.32336351 0.4121245
## bin                  0.07743537 0.1503382
## base_avg_bins_treat 0.31386043 0.4335531
```

Including the -999 street affects the estimate and standard error slightly, but not by much. We'll continue with -999 included, since that's what the original authors did.

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.

The street level is not a strong predictor of recycling.

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[, nocell := xor(havecell, 1)]
```

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.

```
m = d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat + nocell + factor(street))]
coef(summary(m))[1:4,]
```

```
##                         Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)           0.38749355 0.03225539 12.013296  6.836102e-32
## bin                   0.11510074 0.01704496  6.752774  2.023950e-11
## base_avg_bins_treat   0.37338230 0.01429941 26.111720 1.805396e-125
## nocellTRUE           -0.04950989 0.01686377 -2.935873  3.373650e-03
```

```
coefci(m)[1:4,]
```

```
##                            2.5 %      97.5 %
## (Intercept)           0.32422622  0.45076087
## bin                   0.08166792  0.14853357
## base_avg_bins_treat   0.34533472  0.40142988
## nocellTRUE           -0.08258732 -0.01643245
```

Not having a cell phone is not a strong predictor of recycling.

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.

```
m = d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat + nocell + sms + factor(street))]
coef(summary(m))[1:5,]
```

```
##                          Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)          0.384642314 0.03428819 11.217925  3.648065e-28
## bin                  0.115053649 0.01705105  6.747600  2.095786e-11
## base_avg_bins_treat  0.373482860 0.01430947 26.100390 2.285610e-125
## nocellTRUE          -0.046702054 0.02037507 -2.292118  2.202841e-02
## sms                  0.005124375 0.02085563  0.245707  8.059407e-01
```

```
coefci(m)[1:5,]
```

```
##                           2.5 %       97.5 %
## (Intercept)          0.31738773  0.451896896
## bin                  0.08160886  0.148498434
## base_avg_bins_treat  0.34541553  0.401550187
## nocellTRUE          -0.08666674 -0.006737369
## sms                 -0.03578292  0.046031671
```

SMS treatment does not have a strong effect on recycling.

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.

```
m = robust_model(
    d[, lm(avg_bins_treat ~ bin_s + bin_g + sms_p + sms_g + base_avg_bins_treat + nocell + factor(stree
    d$street)
coef(summary(m))[1:7,]
```

```
##                          Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)          0.384943316 0.03429444 11.2246569  3.409256e-28
## bin_s                0.127812892 0.02223628  5.7479441  1.080308e-08
## bin_g                0.103190216 0.02188886  4.7142810  2.636923e-06
## sms_p               -0.008041152 0.02503693 -0.3211716  7.481224e-01
## sms_g                0.019707117 0.02519764  0.7821016  4.342709e-01
## base_avg_bins_treat  0.373852178 0.01431366 26.1185536 1.733949e-125
## nocellTRUE          -0.046383459 0.02037843 -2.2761053  2.297176e-02
```

```
coefci(m, vcov. = m$cluster.vcov)[1:7,]
```

```
##                           2.5 %      97.5 %
## (Intercept)          0.33174771 0.438138919
## bin_s                0.08327698 0.172348803
## bin_g                0.05391168 0.152468752
## sms_p               -0.06653887 0.050456563
## sms_g               -0.03826649 0.077680728
## base_avg_bins_treat  0.31387859 0.433825762
## nocellTRUE          -0.09638673 0.003619815
```

In part g, we did not differentiate between the sticker and non-sticker bins. Therefore, it cannot be compared.

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")
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
```

```
d = data.table(d)
```

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?

```
m = d[, lm(vomiting_day14 ~ treat_zmapp)]
(co = coef(summary(m)))
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  0.8474576 0.05483098 15.455817 5.032102e-28
## treat_zmapp -0.2377015 0.08563161 -2.775862 6.595412e-03
```

The estimated effect is -0.238 (0.086)**. The p-value is** 0.007.

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.

```
m = d[, lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0)]
(co = coef(summary(m)))
```

```
##                      Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)      -19.46965517 7.44094546 -2.6165566 0.010318288
## treat_zmapp       -0.16553674 0.07567142 -2.1875730 0.031128516
## vomiting_day0       0.06455724 0.14635485  0.4411008 0.660131840
## temperature_day0    0.20554815 0.07634039  2.6925215 0.008368374
```

The estimated effect is -0.166 (0.076)**. The p-value is** 0.031.

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

The ATE in part b is a better estimate because it includes pre-treatment covariates that are good predictors of the post-treatment outcome variable.

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.

```
m = d[, lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + temperature_day14)]
(co = coef(summary(m)))
```

```
##                       Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)       -22.59158542 7.47727088 -3.0213678 0.003233343
## treat_zmapp        -0.12010063 0.07767979 -1.5460988 0.125405588
## vomiting_day0       0.04603820 0.14426358  0.3191256 0.750331936
## temperature_day0    0.17664160 0.07641671  2.3115571 0.022963383
## temperature_day14   0.06014826 0.02937296  2.0477427 0.043345384
```

The estimated effect is -0.12 (0.078)**. The p-value is** 0.125.

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

Part b is the better estimate because post-treatment temperature is a bad control. The drug likely has an effect on temperature.

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?

```
m = d[, lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + treat_zmapp*male)]
(co = coef(summary(m)))
```

```
##                      Estimate Std. Error     t value     Pr(>|t|)
## (Intercept)      48.71268983 9.26617942   5.2570415 9.136512e-07
## treat_zmapp      -0.23086555 0.11871003  -1.9447856 5.478966e-02
## vomiting_day0     0.04113066 0.18207655   0.2258976 8.217715e-01
## temperature_day0  0.50479728 0.09508023   5.3091719 7.336866e-07
## male              3.08548611 0.12643666  24.4034141 1.863997e-42
## treat_zmapp:male -2.07668626 0.19163872 -10.8364651 3.109157e-18
```

The coefficient estimate on the interaction term is negative, with a very low p-value. Therefore, the drug is more likely to reduce men's temperatures.

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.)

The colleague performed 20 million regressions. By random chance, 1 million of those regressions should have shown statistical significane at the 0.05 level. A single statistically significant effect cannot be believed, unless the p-value was divided by 20 million after the regression.

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?

We would be more inclined to believe this treatment effect because we are testing specifically for that effect. We did not happen to find it in a list of 20 million other effects.

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.*)

Likely, no experiment can be conducted to test this claim. African ancestry cannot be randomly assigned to other-wise identical groups of people. The question is fundamentally unanswerable.