

# W241 Problem Set 3

Ray Buhr

June 18, 2015

## Problem 1: Effects of Facebook ads

**Skim Broockman and Greens paper on the effects of Facebook ads and download an anonymized version of the data for Facebook users only.**

Read in data:

```
fbads <- read.csv("~/MIDS/241/broockman_green_anon_pooled_fb_users_only.csv", string
sAsFactors = F)
```

**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.**

*Note: Ignore the blocking the article mentions throughout this problem.*

```
study1 <- fbads[fbads$studyno==1,]
s1_fit <- lm(name_recall ~ treat_ad, data = study1)
confint(s1_fit, level = 0.95)
```

```
##                 2.5 %    97.5 %
## (Intercept)  0.15080247 0.21413492
## treat_ad     -0.05101765 0.03142188
```

**b. What are the clusters in Broockman and Greens study? Why might taking clustering into account increase the standard errors?**

The clusters in the study refer to constituents – people with the roughly the same age, gender and geographic location. Taking clustering into account increases the standard error by reducing common information into the smallest group possible. In simple terms, clustering combines information from groups that are similar to reduce the effects of added data points to the sample that don't contribute new information. Reducing the number of samples by clustering usually increases the standard error as the calculation for standard error has sample size in the denominator, so a smaller number would make the calculation larger.

**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.**

*Note: Heres a copy of a function for computing clustered standard errors in regression. Invoke it as follows.*

- my.lm <- lm(x~y, data) # you're accustomed to doing this
- cl(my.lm, data\$cluster) # cluster is a factor variable in the data frame that identifies the clusters

```

cl <- function(fm, cluster) {
  require(sandwich, quietly = TRUE)
  require(lmtest, quietly = TRUE)
  M <- length(unique(cluster))
  N <- length(cluster)
  K <- fm$rank
  dfc <- (M / (M - 1)) * ((N - 1) / (N - K))
  uj <- apply(estfun(fm), 2, function(x) tapply(x, cluster, sum));
  vcovCL <- dfc * sandwich(fm, meat=crossprod(uj)/N)
  coeftest(fm, vcovCL)
}

```

*Hint: to get in order to get the clustering function to work correctly:*

- Remove missing data from the data frame first.
- Turn the cluster variable into a factor with factor(). If you read in the variable as a factor originally (did not set stringsAsFactors as FALSE), you might need to do something like factor(as.character(data\$cluster)) to turn it into a character first then re-turn it into a factor.

```
cl(s1_fit, study1$cluster)
```

```

## 
## Attaching package: 'zoo'
## 
## The following objects are masked from 'package:base':
## 
##     as.Date, as.Date.numeric

```

```

## 
## 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

```

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

```

study2 <- fbads[fbads$studyno==2,]
study2 <- study2[complete.cases(study2),]
s2_fit <- lm(name_recall ~ treat_ad, data = study2)
cl(s2_fit, study2$cluster)

```

```

## 
## 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. 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?**

```

fbads_complete <- fbads[complete.cases(fbads),]
fbads_fit <- lm(name_recall ~ treat_ad, data = fbads_complete)
cl(fbads_fit, fbads_complete$cluster)

```

```

## 
## 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

```

The treatment effect is -0.155 with an associated p-value of 7.344e-09, which is highly significant.

**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?**

```

fbads_complete$study <- fbads_complete$studyno-1
fbads_clustered_fit <- lm(name_recall ~ treat_ad + study, data = fbads_complete)
cl(fbads_clustered_fit, fbads_complete$cluster)

```

```

## 
## 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 
## study       0.4260988 0.0206970 20.5875 <2e-16 *** 
## --- 
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The treatment effect estimate is now -0.0067 with an associated p-value of 0.74, which is not highly significant. The treatment effect of study 2 is 0.426 with an associated p-value of <2e-16, which is highly significant.

**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.)

In Study 1, 60% of subjects received the treatment. In Study 2, only 24% of subjects received the treatment. Analyzing either study by itself leads to biased results because the probability of assignment to the treatment group was not equal across the studies.

**h. Skim this Facebook case study and consider two claims they make reprinted below. Why might their results differ from Broockman and Greens? 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.”

One reason Facebook might find that votes in areas where ads ran to be different than areas where ads did not run could be due to the type of election that existed in those areas. In the Boockman and Green paper, they mention “that online advertisements could prove more effective in high salience contests” (p. 19), which is to suggest that perhaps elections where the candidates are well known the ads become more useful.

Depending on the voting areas chosen by Facebook, the results could be biased due to unforeseen confounding factors, such as preexisting knowledge of candidates or overall more active voters. Subjects who were administered the treatment of Facebook ads were on average exposed to the ad 38 times (p. 18). It could be very possible that repeatedly showing the same ad to a person could cause them to become irritated with the candidate and instead of siding with that person choose to vote against in spite. This idea may help explain why subjects exposed to fewer ads were more likely to vote for the candidate than subjects who received the most ad exposure – a bell curved relationship where more ad exposure helps to a point and quickly turns to increasingly negative ROI.

## Problem 2: Encouraging Recycling in Peru

**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 ATE of providing a recycling bin on the average weight in kg of recyclables turned in per week was +0.187 kg. The 95% confidence interval would be +/- 2 SE (0.032), in this case 0.123 to 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 ATE of sending a text message reminder on the average weight in kg of recyclables turned in per week was -0.024 kg. The 95% confidence interval would be +/- 2 SE (0.039), in this case -0.102 to 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 that turned in a bin, the averages number of bins turned in per week, average weight in kg of recyclables turned in per week, and average market value of recyclables given per week all had statistically significant effects.

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

No outcome measures showed statistically significant effects of sending text messages.

**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.**

For each 1 kg of recyclables turned in per week the study estimates a +0.281 increase in average weight of recyclables turned in per week. All other things being equal, we would expect an extra 2 kg per week of recyclables turned in prior to treatment to predict 0.562 kg of recyclables turned in during the treatment period.

**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.**

There most likely is some overlap in effect between percentage of visits turned in bag with the average number of bins turned in and average weight of recyclables turned in, so removing that variable would expect to increase the estimated ATE. Without the extra variable, there would be an increase in the standard error since there are fewer other factors to attribute to variance.

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

No, has cell phone is not a bad control because it does not inherently affect recycling bins turned in, though it almost certainly does effect whether an SMS message was received.

**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.**

Removing the has cell phone variable would increase the coefficient of any SMS message since the two variables are almost assuredly positively correlated.

## Problem 3: More Recycling in Peru!

**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 is multifactorial experimental design with 3x3 dimensions. There are two variables, SMS or bin, each with three different variations. For bin, the variations were no bin, generic bin or bin with sticker. For SMS, the variations were no SMS, generic SMS, or personalized SMS.

**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 establishes what effect we would estimate just by chance without any treatment administered. For this study, the baseline results indicate that no bins previously turned in would turn in +0.374 bins, no weight in kg of recyclables would expect to turn in +0.281 bins, no market value for recyclables would expect to turn in +0.233 bins and no percent contamination would expect to turn in +0.292 bins.

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

The coefficient for bin without sticker is +0.035, which means that a subject issued a bin, but one without a sticker, would expect to submit 0.035 more bins than a subject who did not receive a bin.

**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 recycling bin with sticker has a greater coefficient value, 0.055, than without a sticker, 0.035, which is a difference of 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.**

No, this is not statistically significant. The F-test p-value of 0.31, which is not significant. The standard error for that treatment is 0.015, so 2 times the SE is greater than the magnitude of the difference, which would also suggest the difference is not 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 fully saturated model includes indicators for each unique combination of treatments, thus removing the differences in estimated effect. Since each variable is the group of treatment or non-treatment for each variable, the model is inherently linear and thus should represent a perfect fit to the data but provide little statistical value for causation.

## Problem 4: What, More Recycling in Peru?

Download the data set for the recycling study in the previous problem, obtained from the authors. (The data is in Stata format. To load it into R, use library(foreign) then read.dta("datafilename.dta").) Well be focusing on the outcome variable Y=number of bins turned in per week

```
library(foreign)
recycle <- read.dta("~/MIDS/241/karlan_data_subset_for_class.dta")
```

**a. For simplicity, lets 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.**

```
a <- lm(avg_bins_treat ~ bin, data = recycle)
summary(a)
```

```
##
## Call:
## lm(formula = avg_bins_treat ~ bin, data = recycle)
##
## 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
```

```
confint(a)
```

```
##              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.**

```
b <- lm(avg_bins_treat ~ bin + base_avg_bins_treat, data = recycle)
summary(b)
```

```

## 
## Call:
## lm(formula = avg_bins_treat ~ bin + base_avg_bins_treat, data = recycle)
## 
## 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

```

```
confint(b)
```

```

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

```

Adding the pre-treatment value allows us to explain more of the random variation, which decreases standard error and thus provides a tighter confidence interval.

**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.**

```

c <- lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street), data = recycle)
cf_c <- confint(c)
cf_c[1:2,]

```

```

##                2.5 %    97.5 %    
## (Intercept) 0.30573165 0.4297564
## bin         0.08042873 0.1473448

```

**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.**

Confidence interval changed from (0.0920038, 0.1573822) to (0.0804287, 0.1473448) when adding street level blocking. Blocking by street reduces variability in the treatment conditions and potential effects of confounding variables as long as the variability between different streets is greater than the variability within

each street. This seems likely as street is often an indicator of wealth/income, which could potentially be an indicator of recycling behavior and a confounding variable.

**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.**

```
recycle$no_cell_phone <- recycle$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.**

```
f <- lm(avg_bins_treat ~ bin + base_avg_bins_treat + no_cell_phone + factor(street),
  data = recycle)
cf_f <- confint(f)
cf_f[1:4, ]
```

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

The confidence interval changed from (0.0804287, 0.1473448) to (0.0816679, 0.1485336) when adding the covariate no cell phone. This confidence interval is not much different because not having a cell phone does not explain any additional variance in the expected effect on recycling. Perhaps the variability in street and having a cell phone overlap enough that no additional information is gained.

**g. Now lets 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.**

```
g <- lm(avg_bins_treat ~ bin + base_avg_bins_treat + no_cell_phone + sms + factor(street),
  data = recycle)
g$coefficients[1:5]
```

```
##                                (Intercept)          bin   base_avg_bins_treat
## 0.384642314                  0.115053649        0.373482860
## no_cell_phone                      sms
## 0.046702054                  0.005124375
```

```
cf_g <- confint(g)
cf_g[1:5, ]
```

```

##                               2.5 %      97.5 %
## (Intercept)          0.317387732 0.45189690
## bin                  0.081608864 0.14849843
## base_avg_bins_treat 0.345415534 0.40155019
## no_cell_phone        0.006737369 0.08666674
## sms                 -0.035782922 0.04603167

```

SMS did not have much of an effect on the expected bins to recycle, so the confidence interval did not shift much as there was not additional explanation of variance in the model.

**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.**

```

h <- lm(avg_bins_treat ~ bin + bin_s + bin_g +
         sms + sms_p + sms_g + no_cell_phone +
         factor(street), data = recycle[complete.cases(recycle),])
h$coefficients[1:8]

```

	bin	bin_s	bin_g	sms
## (Intercept)	0.1158250804	0.0112733048	NA	0.0001667005
## sms_p	0.6681553299	0.0580883175	NA	-0.0205114374

```
cf_h <- confint(h)
```

By adding in variations on treatment effects, the study added more confounding variables which increased the standard error and thus decreased the ability for the study to show the effects on recycling.

## Problem 5: Experimental Drug ZMapp

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. 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.

```
zmapp_data <- read.csv("~/MIDS/241/ebola_rct2.csv", stringsAsFactors = F)
```

**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?**

```

a <- lm(vomiting_day14 ~ treat_zmapp, data = zmapp_data)
summary(a)

```

```

## 
## Call:
## lm(formula = vomiting_day14 ~ treat_zmapp, data = zmapp_data)
## 
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -0.84746 -0.03803  0.15254  0.21197  0.39024 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  0.84746   0.05483 15.456 <2e-16 ***  
## treat_zmapp -0.23770   0.08563 -2.776  0.0066 **  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.4212 on 98 degrees of freedom
## Multiple R-squared:  0.0729, Adjusted R-squared:  0.06343 
## F-statistic: 7.705 on 1 and 98 DF, p-value: 0.006595

```

The estimated effect of ZMapp on whether someone was vomiting on day 14 is -0.2377 (0.08563). The associated p-value is 0.0066, which is statistically significant at the 95% level.

**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.**

```

b <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0, data = zmapp_data)
summary(b)

```

```

## 
## Call:
## lm(formula = vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
##      data = zmapp_data)
## 
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -0.79643 -0.18106  0.04654  0.23122  0.68413 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -19.46966   7.44095 -2.617  0.01032 *  
## treat_zmapp  -0.16554   0.07567 -2.188  0.03113 *  
## vomiting_day0  0.06456   0.14635  0.441  0.66013  
## temperature_day0 0.20555   0.07634  2.693  0.00837 ** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.3668 on 96 degrees of freedom
## Multiple R-squared:  0.311, Adjusted R-squared:  0.2895 
## F-statistic: 14.45 on 3 and 96 DF, p-value: 7.684e-08

```

The ATE for ZMapp was -0.16554 (0.07567). The associated p-value was 0.03113, which is still significant at the 95% level.

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

The confidence interval shrank dramatically after adding the covariates in part b, from (-0.4076347) to (-0.3157433, -0.0153302). I prefer the result in part b because it has a much higher R-squared value and F-statistic so is probably more accurate.

**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.**

```
d <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + temperature_day14, data = zmapp_data)
summary(d)
```

```
##
## Call:
## lm(formula = vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 +
##     temperature_day14, data = zmapp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87745 -0.27436  0.04701  0.24801  0.66445
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -22.59159    7.47727 -3.021  0.00323 ***
## treat_zmapp -0.12010    0.07768 -1.546  0.12541
## vomiting_day0  0.04604    0.14426  0.319  0.75033
## temperature_day0  0.17664    0.07642  2.312  0.02296 *
## temperature_day14  0.06015    0.02937  2.048  0.04335 *
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3609 on 95 degrees of freedom
## Multiple R-squared:  0.3402, Adjusted R-squared:  0.3124
## F-statistic: 12.24 on 4 and 95 DF,  p-value: 4.545e-08
```

The ATE after adding covariate for temperature on day 14 was -0.1201 (0.07768), with an associated p-value of 0.12541.

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

```
cf_b <- confint(b)
cf_b[2,]
```

```
##      2.5 %    97.5 %
## -0.31574331 -0.01533017
```

```
cf_d <- confint(d)
cf_d[2,]
```

```
##      2.5 %    97.5 %
## -0.27431451  0.03411325
```

Comparing the confidence intervals between b and d we can see that adding temperature on day 14 shifted the range of expected treatment effect, but did not drastically change the magnitude of the range. That said, the F-statistic did decrease so the model in d is less predicted of the total variation than the model in b.

**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?**

```
f_male <- lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
               data = zmapp_data[zmapp_data$male==1,])
f_female <- lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
               data = zmapp_data[zmapp_data$male!=1,])
summary(f_male)
```

```
##
## Call:
## lm(formula = temperature_day14 ~ treat_zmapp + vomiting_day0 +
##     temperature_day0, data = zmapp_data[zmapp_data$male == 1,
##     ])
##
## Residuals:
##      Min        1Q        Median         3Q        Max
## -0.75330 -0.37437 -0.01924  0.28514  0.82581
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.2316   15.6982   1.671   0.104
## treat_zmapp -2.2391    0.1538 -14.559 6.38e-16 ***
## vomiting_day0 -0.3791    0.2972  -1.276   0.211
## temperature_day0  0.7670    0.1609   4.766 3.67e-05 ***
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4399 on 33 degrees of freedom
## Multiple R-squared:  0.9178, Adjusted R-squared:  0.9104
## F-statistic: 122.9 on 3 and 33 DF,  p-value: < 2.2e-16
```

```
summary(f_female)
```

```

## 
## Call:
## lm(formula = temperature_day14 ~ treat_zmapp + vomiting_day0 +
##      temperature_day0, data = zmapp_data[zmapp_data$male != 1,
##      ])
## 
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -0.77864 -0.38407 -0.05783  0.39070  0.76345 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 61.4633   11.3406   5.420 1.16e-06 *** 
## treat_zmapp -0.2291    0.1189  -1.927  0.05881 .    
## vomiting_day0  0.2655    0.2269   1.170  0.24665    
## temperature_day0  0.3739    0.1164   3.212  0.00214 **  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.4516 on 59 degrees of freedom 
## Multiple R-squared:  0.5543, Adjusted R-squared:  0.5317 
## F-statistic: 24.46 on 3 and 59 DF,  p-value: 2.073e-10

```

To compare the hypothesis that male reduces temperature, you would want to block on the male variable. Running the regression on only male participants shows a coefficient of -2.239, compared to non-male coefficient of -0.229. This evidence seems to support the notion that ZMapp is especially likely to reduce men's temperatures compared to women's.

**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 treatments 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.)**

Since the study included more women than men, and a higher percentage of men were treated than women, it is hard to say for certain that the ZMapp drug actually works better on men.

**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?**

No more likely to believe because experimental results cannot infer the covariance between  $Y(0)$  and  $Y(1)$ .

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*.)*

The ability to change race is far-fetched so this is a fundamentally unanswerable question.