

Problem Set 3

Experiments and Causality

Adam Yang

```
# 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)
```

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.

```
# Equation to calculate robust standard errors
processrobust <- function(linReg, row, z_val){
  # First create a summary of the linReg for confidence interval
  model <- summary(linReg)
  # Get Variance-Covariance matrix
  vcov.robust <- vcovHC(linReg)
  # Calculate standard errors by taking the sqrt of the diagonals of vcov
  robust.se <- sqrt(diag(vcov.robust))
  # Calculate the 95% confidence interval using cluster se
  slope <- model$coefficients[row,1]
  l_bound <- signif(slope - z_val*robust.se, 3)
  u_bound <- signif(slope + z_val*robust.se, 3)

  # return lower bound, upper bound, cluster se
  return(c(l_bound[row], u_bound[row], signif(robust.se[row],3)))
}

# Equation to calculate clustered standard errors and confidence interval
processcluster <- function(linReg, row, z_val) {
  # First create a summary of the linReg for confidence interval
  model <- summary(linReg)
  # Get clustered Variance-Covariance matrix
```

```
vcov.cluster <- cluster.vcov(linReg, ~ cluster)
# Calculate standard errors by taking the sqrt of the diagonals of vcov
cluster.se <- sqrt(diag(vcov.cluster))
# Calculate the 95% confidence interval using cluster se
slope <- model$coefficients[row,1]
l_bound <- signif(slope - z_val*cluster.se, 3)
u_bound <- signif(slope + z_val*cluster.se, 3)

# return lower bound, upper bound, cluster se
return(c(l_bound[row], u_bound[row], signif(cluster.se[row],3)))
}

# Equation to calculate the default confidence interval
conf.interval <- function(linReg, row, z_val) {
  # 95% confidence interval = Estimate +/- 1.96(Std. Error)  slope <- model$coefficients[row,1]
  model <- summary(linReg)
  std.err <- model$coefficients[row,2]
  slope <- model$coefficients[row,1]
  l_bound <- signif(slope - z_val*std.err, 3)
  u_bound <- signif(slope + z_val*std.err, 3)

  return(c(l_bound, u_bound))
}
```

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")
```

- 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.

```
Study1 <- d[d$studyno == 1,]
```

```
# Use regression without clustered standard errors
model1.a <- lm(name_recall~treat_ad, data = Study1)
```

```
# Compute a confidence interval for the effect of the ad on name_recall
c_int <- conf.interval(model1.a, row = 2, z_val = 1.96)
```

```
# Compute a robust confidence interval for the effect of the ad on name_recall
ans <- processrobust(model1.a, row = 2, z_val = 1.96)
```

```
paste("The 95% confidence interval on only Study 1 w/o clustering is (", c_int[1],",",c_int[2],")")
```

```
## [1] "The 95% confidence interval on only Study 1 w/o clustering is ( -0.051 , 0.0314 )"
```

```
paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0211"
paste("The robust 95% confidence interval on only Study 1 w/o clustering is (", ans[1], ",", ans[2], ")")
```

```
## [1] "The robust 95% confidence interval on only Study 1 w/o clustering is ( -0.0512 , 0.0316 )"
```

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

Adam Yang: In Broockman and Green's study, they performed clustering by age (or range of ages), gender, and location. For example, a specific cluster might be 24 year old males from New York. The clustered standard errors are larger when there is more variability of the cluster level means. Usually the average potential outcomes of each cluster can have high variability and therefore give us large standard errors based on the clustered standard error equation. However, if the cluster level potential outcomes have very small variability, it is possible for the clustered standard error to not be as big.

- 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`.

```
# Create regression
model1.c <- lm(name_recall~treat_ad, data = Study1)
# Put regression through function to get lbound, ubound, cluster.se
ans <- processcluster(model1.c, row = 2, z_val = 1.96)

paste("The clustered standard error for only Study 1 is", ans[3])

## [1] "The clustered standard error for only Study 1 is 0.0238"
paste("The 95% confidence interval on only Study 1 w/ clustering is (", ans[1], ",", ans[2], ")")

## [1] "The 95% confidence interval on only Study 1 w/ clustering is ( -0.0564 , 0.0368 )"
```

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

```
Study2 <- d[d$studyno == 2,]

# Create regression
model1.d <- lm(name_recall~treat_ad, data = Study2)
# Put regression through function to get lbound, ubound, cluster.se
ans <- processcluster(model1.d, row = 2, z_val = 1.96)

paste("The clustered standard error is", ans[3])

## [1] "The clustered standard error is 0.0355"
paste("The 95% confidence interval on only Study 2 w/ clustering is (", ans[1], ",", ans[2], ")")

## [1] "The 95% confidence interval on only Study 2 w/ clustering is ( -0.0724 , 0.0668 )"
```

- 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?

Note To Grader: I am assuming that for this question, we are still clustering by `cluster` column. The `cluster` column differentiates between "study 1 cluster 1" and "study 2 cluster 1". I am assuming that "study

1 cluster 1” and “study 2 cluster 1” are 2 very different clusters and therefore I don’t have to create a new column with that removes the study number from the cluster names.

```
# Create regression
model1.e <- lm(name_recall~treat_ad, data = d)
# Put regression through function to get lbound, ubound, cluster.se
ans <- processcluster(model1.e, row = 2, z_val = 1.96)
# Use coeftest to find the cluster-robust p value
coeftestvalues <- coeftest(model1.e, cluster.vcov(model1.e, ~ cluster))

paste("The clustered standard error is", ans[3])

## [1] "The clustered standard error is 0.0267"
paste("The 95% confidence interval on the entire sample w/ clustering is (", ans[1],",",ans[2],",)")

## [1] "The 95% confidence interval on the entire sample w/ clustering is ( -0.207 , -0.103 )"
paste("The treatment effect estimate is", signif(summary(model1.e)$coefficients[2,1],3))

## [1] "The treatment effect estimate is -0.155"
# I don't think the following line is necessary but I kept it in anyways
paste("The associated p-value is", signif(summary(model1.e)$coefficients[2,4],3))

## [1] "The associated p-value is 2.16e-16"
paste("The associated cluster-robust p-value is", signif(coeftestvalues[2,4],3))

## [1] "The associated cluster-robust p-value is 7.34e-09"
```

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?

```
# Create new dataframe but create a dummy variable for study number
data1f <- d
data1f$isStudy2 <- ifelse(data1f$studyno == 2, 1, 0)

# Create regression, but this time include dummy variable
model1.f <- lm(name_recall~treat_ad + isStudy2, data = data1f)

# Put regression through function to get lbound, ubound, cluster.se
ans <- processcluster(model1.f, row = 2, z_val = 1.96)

# Use coeftest to find the cluster-robust p value
coeftestvalues <- coeftest(model1.f, cluster.vcov(model1.f, ~ cluster))

paste("The clustered standard error is", ans[3])

## [1] "The clustered standard error is 0.0204"
paste("The 95% confidence interval on the entire sample w/ clustering is (", ans[1],",",ans[2],",)")

## [1] "The 95% confidence interval on the entire sample w/ clustering is ( -0.0468 , 0.0332 )"
paste("The treatment effect estimate with dummy variable is", signif(summary(model1.f)$coefficients[2,1],1))

## [1] "The treatment effect estimate with dummy variable is -0.00678"
# I don't think the following line is necessary but I kept it in anyways
paste("The associated p-value is", signif(summary(model1.f)$coefficients[2,4],3))
```

```
## [1] "The associated p-value is 0.709"
```

```
paste("The associated cluster-robust p-value is", signif(coeftestvalues[2,4],3))
```

```
## [1] "The associated cluster-robust p-value is 0.74"
```

- 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.)

Adam Yang: The two results differ because one of them does not take into account which study the data is from and therefore, completely ignoring the differences between the two studies. The result in part (e) is biased because it does not take into consideration how study 1 and study 2 might have different treatment assignment rates. In other words, when the probability of being assigned to the treatment group varies by block between study 1 and study 2, you can get a biased estimate of the ATE if you don't take this into consideration in your regression. Part (f), on the other hand, introduces a dummy variable as a covariate which helps take account of the impact of the data being from study 2 rather than study 1.

- 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."

Adam Yang: The results of the two experiments may differ because of the many differences between the two experiments. One major difference is that Chong and Koster studied the effects on the "Vote no on 8 in Florida" campaign while Broockman and Green studied the effects on a Republican state legislative candidate in a non-battleground state as well as the effects on a Democratic congressional candidate in a non-battleground state. In one case we are studying the treatment effect on an amendment, while in the other case, we are studying the treatment effect on candidates running for certain offices. People can possibly have very different reactions to Facebook ads based on amendments versus candidates. Also, Chong and Koster's experiment took place in 2010 while Broockman and Green's experiment took place in 2012. Facebook is a rapidly growing social media platform so the time difference between 2010 and 2012 might have a huge impact on the UI of Facebook, people's attitude towards ads on Facebook, the number and type of people who have a Facebook account, among many other different possible differences. Also, in Chong and Koster's experiment, they focused on two of the most populated counties in Florida to do their experiment. It is not clear which non-battleground states Broockman and Green carried out their experiments but the treatment effect on these different locations can be very different. There may be many other reasons why the two experiments are different, but in summary, you cannot compare the results from the two experiments because it would not be an apples to apples comparison.

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.

Adam Yang: The estimated ATE of providing a recycling bin on the average weight of recyclables turned in per household per week is **0.187kg**. The 95% confidence interval can be calculated by $coef \pm 1.96 * se =$

$0.187 \pm 1.96 * 0.032$. Therefore, the 95% confidence interval is between **0.124kg** and **0.250kg**. This result is statistically significant according to the table.

- 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.

Adam Yang: The estimated ATE of sending a text message reminder on the average weight of recyclables turned in per household per week is **-0.024kg**. The 95% confidence interval can be calculated by $coef \pm 1.96 * se = -0.024 \pm 1.96 * 0.039$. Therefore, the 95% confidence interval is between **-0.100kg** and **0.0524kg**. According to the table, this result is not statistically significant.

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

Adam Yang: The outcome measures of: percentage of visits turned in bag, average 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. In other words, only the outcome measure shown in column 5 did not show a statistically significant effect of providing recycling bin.

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

Adam Yang: None of the outcome measures in Table 4A show 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.

Adam Yang: The baseline coefficient for the average weight of recyclables turned in per week is 0.281. That means that for every 1kg of baseline recyclable that a household turns in, they will experience an additional 0.281kg of recycling during the treatment weeks. Therefore, we predict household A would have $2 * 0.281 = 0.562kg$ more recycling than household B per week during the treatment weeks.

- 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.

Adam Yang: If the variable “percentage of visits turned in bag, baseline” had been left out of the regression, the resulting coefficients of each of your variables are expected to change though it is not necessarily clear which direction they will change. It seems that the baseline is positively correlated to the outcome variable. If providing a recycling bin causes an increase in the baseline, then we can expect an increase in the estimated ATE on providing a recycling bin. However, I am not sure if this is the case so I am not sure of the direction of the change. It is also possible that there would be little or no change to the estimated ATE. The standard error will go up because the variance in the measurements captured from this variable will be re-absorbed into our error term.

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

Adam Yang: Because “has cell phone” is not a post treatment variable, I would not say it is a bad control. In fact, I think it is a covariate that is important to include because one of our treatment variables is whether or not participants received texts to remind them to recycle. I cannot think of a reason for how the outcome variable would cause the participant to own a cellphone or not. Furthermore, the status of having a cellphone has been obtained before the experiment was conducted. “has cell phone” is not a statically significant variable, but that does not make it a bad control.

- 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.

Adam Yang: If we remove the “has cell phone” variable from the regression, then we would introduce an omitted variable bias to the coefficient on “Any SMS message”. If you receive an SMS message, you are likely to have a cell phone so “has cell phone” is positively correlated with the “Any SMS message” variable. Furthermore, according to the paper, “cell phone owners in the participation study are slightly richer, more educated, and more interested in local affairs (especially recycling matters) than non-cell phone owners”. This means that “has cell phone” is positively correlated with recycling participation intensity as well. Therefore, the omitted variable bias would’ve made the coefficient of “Any SMS message” larger than shown in Table 4A.

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.)

Adam Yang: The participation intensity experiment uses a sample frame of households that have already recycled through PRISMA previously. The participants are randomly assigned to receiving a recycling bin with a sticker, a recycling bin without a sticker, and no recycling bin at all. Furthermore, the participants are also randomly assigned to receive weekly personalized text message reminders, weekly generic text message reminders, and no text message at all.

This is a 3x3 experiment because there are three groups for recycling bin and 3 groups for text messages. The groups for recycling bin are: bin provided with sticker with recycling information, bin provided with no sticker, and no bin (control). The groups for text messages are: personal SMS message reminding the person to recycle, generic SMS message reminding person to recycle, and no text message (control). TO DO: 3x4x5?

- 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?

Adam Yang: In table 4b, we have 5 different baseline categories. Each of these baseline categories store the baseline recycling intensity (measured differently in each baseline category) of the households prior to treatment. The attributes of the group of people for whom all dummy variables are equal to zero include not being given a recycling bin, not receiving SMS text reminders, and not owning a cell phone. The regression of this group of people can be predicted from only the baseline term.

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

Adam Yang: The coefficient in column 1 on “bin without sticker” is 0.035. This means that, holding everything else constant, by giving a household a recycling bin without a sticker, you can expect the percentage of visits where the household turned in a bag to increase by 0.035 or 3.5%.

- 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?

Adam Yang: The ATE of recycling bin with message sticker is 0.055 while the recycling bin without a sticker is 0.035. Both are statistically significant. It seems like recycling bin with message sticker has a stronger treatment effect. The magnitude of the estimated difference is 0.020.

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

Adam Yang: The F-test p-value (1) = (2) is 0.31 which is not statistically significant. This means that we cannot reject the null hypothesis that the difference between coefficient (1) and coefficient (2) is 0.

- 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.”

Adam Yang: A fully saturated model means that each value of the covariate gets a different dummy variable to represent it. Furthermore, all the possible interactions between these dummy variables are also included in the regression model. In terms of table 4C, the model is saturated in regards to having all the interaction terms the three different types of bin groups. You can see interaction terms between the recycling bin dummy variables with each of the three SMS message dummy variables as well as the phone dummy variables. It is possible that they included the dummy variables themselves (ie. “Bin with sticker”) and all the other interaction terms, but they are not shown in table 4C. They did mention that they had “indicators for each unique combination of treatments”, underneath table 4C.

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
```

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

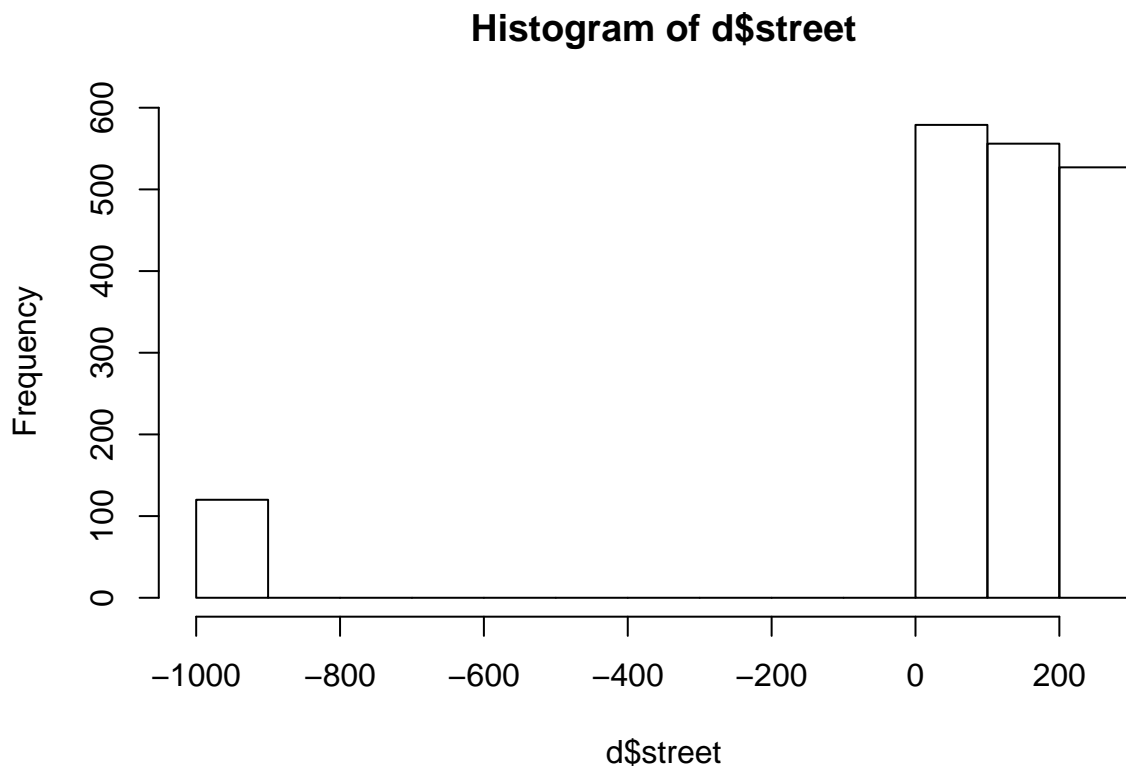
```
summary(d)
```

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



```
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000
##
## sms_p sms_g
## Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000
## Mean :0.1557 Mean :0.1529
## 3rd Qu.:0.0000 3rd Qu.:0.0000
## Max. :1.0000 Max. :1.0000
##
```

```
# Histogram of street
hist(d$street)
```



```
# Figure out how many rows have -999 for street
paste("There are", nrow(d[d$street == -999,]), "rows in the `street` column with -999")
```

```
## [1] "There are 123 rows in the `street` column with -999"
```

Adam Yang: Looking at the summary of the data, we notice that `street` and `havecell` have NA values while the other variables do not. Furthermore, `street` seems to be represented by integer ids, however, there are quite a few of them with values -999. I assume these are some kind of error output values written in the code that produces this data set. If we leave it in the data set, we are assuming all 123 of those data points have the same street, which doesn't make sense. I decided to remove all the rows with NA in `street` and `havecell` as well as all the rows with -999 in `street`.

```
# Clean up the data set by removing rows with NA's and -999 from street and havecell
cleaned_d <- d[!is.na(d$havecell) & !is.na(d$street) & d$street != -999,]
```

- 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.

Note to Grader: I am using the robust standard error for each of the calculations below.

```
model4a <- lm(avg_bins_treat ~ bin, data = cleaned_d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model4a, row = 2, z_val = 1.96)

summary(model4a)

##
## Call:
## lm(formula = avg_bins_treat ~ bin, data = cleaned_d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7698 -0.2602 -0.0518  0.1885  3.5315
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.63515     0.01223   51.924 < 2e-16 ***
## bin          0.13470     0.02086    6.456 1.41e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4039 on 1659 degrees of freedom
## Multiple R-squared:  0.02451,    Adjusted R-squared:  0.02392
## F-statistic: 41.68 on 1 and 1659 DF,  p-value: 1.406e-10

paste("The treatment effect estimate is", signif(summary(model4a)$coefficients[2,1],3))

## [1] "The treatment effect estimate is 0.135"

paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0213"

paste("The 95% robust confidence interval is (", ans[1], ",", ans[2], ")")

## [1] "The 95% robust confidence interval is ( 0.0929 , 0.177 )"
```

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.

```
model4b <- lm(avg_bins_treat ~ bin + base_avg_bins_treat, data = cleaned_d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model4b, row = 2, z_val = 1.96)

summary(model4b)

##
## Call:
## lm(formula = avg_bins_treat ~ bin + base_avg_bins_treat, data = cleaned_d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.97116 -0.21375 -0.02318 0.16776 2.13259
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.35441    0.01425  24.868 < 2e-16 ***
## bin            0.12711    0.01722   7.383 2.44e-13 ***
## base_avg_bins_treat 0.38535    0.01381  27.905 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3332 on 1658 degrees of freedom
## Multiple R-squared:  0.3362, Adjusted R-squared:  0.3354
## F-statistic: 419.9 on 2 and 1658 DF, p-value: < 2.2e-16

paste("The treatment effect estimate is", signif(summary(model4b)$coefficients[2,1],3))

## [1] "The treatment effect estimate is 0.127"

paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0177"

paste("The 95% robust confidence interval is (", ans[1],",",ans[2],")")

## [1] "The 95% robust confidence interval is ( 0.0923 , 0.162 )"
```

Adam Yang: From parts (a) and (b) we see that the confidence interval is slightly tighter when we introduce the pre-treatment value of Y as a covariate. The treatment effect also got smaller. The pre-treatment value of the Y as a covariate helps us better understand the variation of the data and therefore decreasing our robust standard error from 0.0213 to 0.0177. Therefore the confidence interval got tighter.

- 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.

```
model4c <- lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street), data = cleaned_d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model4c, row = 2, z_val = 1.96)

summary(model4c)

##
## Call:
## lm(formula = avg_bins_treat ~ bin + base_avg_bins_treat + factor(street),
##     data = cleaned_d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42645 -0.19281 -0.01686  0.16072  1.86953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2761452  0.0979261   2.820 0.004867 **
## bin            0.1160641  0.0176099   6.591 6.07e-11 ***
## base_avg_bins_treat 0.3669578  0.0147709  24.843 < 2e-16 ***
## factor(street)3    0.0505660  0.1413142   0.358 0.720524
## factor(street)4   -0.0413983  0.1453562  -0.285 0.775833
## factor(street)5    0.0011216  0.1453437   0.008 0.993844
```

## factor(street)6	0.0442730	0.1503140	0.295	0.768389	
## factor(street)7	0.0388734	0.1744273	0.223	0.823673	
## factor(street)8	0.0407896	0.1302853	0.313	0.754264	
## factor(street)9	0.0082076	0.1349771	0.061	0.951521	
## factor(street)10	0.0679267	0.1225271	0.554	0.579402	
## factor(street)11	0.0272125	0.1352207	0.201	0.840535	
## factor(street)15	0.0741247	0.1889864	0.392	0.694950	
## factor(street)17	0.1349727	0.1641210	0.822	0.410983	
## factor(street)20	-0.1377622	0.1504508	-0.916	0.359993	
## factor(street)21	0.0177666	0.1641569	0.108	0.913829	
## factor(street)22	-0.0298983	0.1156727	-0.258	0.796078	
## factor(street)23	-0.0638941	0.1177896	-0.542	0.587595	
## factor(street)26	0.0696100	0.2107055	0.330	0.741170	
## factor(street)32	0.2504634	0.2107280	1.189	0.234802	
## factor(street)37	0.2673750	0.1744992	1.532	0.125676	
## factor(street)38	0.0223127	0.1744275	0.128	0.898230	
## factor(street)40	0.0337240	0.2107826	0.160	0.872908	
## factor(street)41	0.1088629	0.1744597	0.624	0.532724	
## factor(street)42	-0.0535698	0.2107958	-0.254	0.799429	
## factor(street)43	0.1480805	0.1744210	0.849	0.396028	
## factor(street)44	0.2291915	0.1503031	1.525	0.127507	
## factor(street)45	0.0551281	0.1888010	0.292	0.770335	
## factor(street)46	0.1070460	0.1641766	0.652	0.514491	
## factor(street)47	0.5185747	0.1891575	2.741	0.006190	**
## factor(street)53	0.1321706	0.2106366	0.627	0.530440	
## factor(street)58	0.0646041	0.1324919	0.488	0.625900	
## factor(street)60	0.1225841	0.1266497	0.968	0.333253	
## factor(street)61	0.0853744	0.2110069	0.405	0.685827	
## factor(street)62	0.1524973	0.1563694	0.975	0.329602	
## factor(street)63	0.0655447	0.1502548	0.436	0.662738	
## factor(street)64	0.1467600	0.1641345	0.894	0.371390	
## factor(street)66	0.0356481	0.1563689	0.228	0.819698	
## factor(street)67	0.0834804	0.1453874	0.574	0.565924	
## factor(street)68	0.1680898	0.1502563	1.119	0.263455	
## factor(street)69	0.1728585	0.1379698	1.253	0.210450	
## factor(street)70	0.0861086	0.1302902	0.661	0.508780	
## factor(street)72	0.0502082	0.1380116	0.364	0.716062	
## factor(street)73	-0.0760441	0.1503173	-0.506	0.613009	
## factor(street)74	0.0291038	0.1213819	0.240	0.810541	
## factor(street)75	0.0377228	0.1349974	0.279	0.779951	
## factor(street)77	0.0609375	0.1564266	0.390	0.696918	
## factor(street)78	0.0426430	0.1266500	0.337	0.736391	
## factor(street)79	0.4020718	0.1888568	2.129	0.033421	*
## factor(street)80	0.1068185	0.1156742	0.923	0.355927	
## factor(street)81	0.0615651	0.1888667	0.326	0.744492	
## factor(street)82	0.1268644	0.2107252	0.602	0.547241	
## factor(street)83	-0.0263435	0.1566875	-0.168	0.866506	
## factor(street)84	0.0768420	0.1563378	0.492	0.623137	
## factor(street)85	0.1595347	0.1266489	1.260	0.207990	
## factor(street)86	0.0295295	0.1226093	0.241	0.809710	
## factor(street)88	-0.0007989	0.1413653	-0.006	0.995492	
## factor(street)89	0.0896628	0.1503132	0.597	0.550928	
## factor(street)91	-0.1177081	0.2106337	-0.559	0.576363	
## factor(street)93	0.0243957	0.1563558	0.156	0.876033	

## factor(street)94	0.0564087	0.1453768	0.388	0.698059	
## factor(street)96	0.0553585	0.1563592	0.354	0.723354	
## factor(street)98	0.1472731	0.1351000	1.090	0.275845	
## factor(street)99	0.3263651	0.1455217	2.243	0.025062	*
## factor(street)100	0.1728747	0.1415513	1.221	0.222172	
## factor(street)101	0.2941672	0.1454326	2.023	0.043283	*
## factor(street)102	0.0192946	0.2117854	0.091	0.927422	
## factor(street)103	0.1799488	0.1453467	1.238	0.215888	
## factor(street)105	0.2129932	0.1745108	1.221	0.222464	
## factor(street)106	0.4959513	0.1350190	3.673	0.000248	***
## factor(street)107	0.1602545	0.1888154	0.849	0.396165	
## factor(street)109	0.0815379	0.1502912	0.543	0.587533	
## factor(street)110	0.2363847	0.1744384	1.355	0.175587	
## factor(street)111	0.0980077	0.1744880	0.562	0.574414	
## factor(street)112	0.2840673	0.1505883	1.886	0.059439	.
## factor(street)113	0.1776071	0.1566009	1.134	0.256920	
## factor(street)115	0.3896666	0.1350929	2.884	0.003978	**
## factor(street)117	0.6636662	0.2122053	3.127	0.001798	**
## factor(street)118	0.1293686	0.2106588	0.614	0.539234	
## factor(street)119	0.0853025	0.1503141	0.567	0.570464	
## factor(street)120	0.0294091	0.1503476	0.196	0.844944	
## factor(street)121	0.2768029	0.1505016	1.839	0.066086	.
## factor(street)122	-0.0637493	0.1888629	-0.338	0.735756	
## factor(street)124	0.1345711	0.1378936	0.976	0.329271	
## factor(street)125	0.3996890	0.1272343	3.141	0.001715	**
## factor(street)126	-0.0040099	0.1645758	-0.024	0.980565	
## factor(street)127	0.1736553	0.1645780	1.055	0.291526	
## factor(street)128	0.2415746	0.1350640	1.789	0.073885	.
## factor(street)129	0.4739654	0.2108342	2.248	0.024720	*
## factor(street)130	0.2732824	0.1503850	1.817	0.069386	.
## factor(street)131	0.3753228	0.1507911	2.489	0.012919	*
## factor(street)132	0.3331701	0.1286646	2.589	0.009707	**
## factor(street)133	0.4114377	0.1888639	2.178	0.029527	*
## factor(street)134	0.0869171	0.1565116	0.555	0.578746	
## factor(street)136	0.2799097	0.1505580	1.859	0.063204	.
## factor(street)137	0.3515908	0.2106588	1.669	0.095328	.
## factor(street)138	0.2200732	0.1384773	1.589	0.112221	
## factor(street)147	0.2879270	0.2107472	1.366	0.172078	
## factor(street)148	0.2019212	0.1888154	1.069	0.285059	
## factor(street)149	0.3798725	0.1457037	2.607	0.009221	**
## factor(street)151	-0.1638204	0.2109091	-0.777	0.437439	
## factor(street)152	0.2197070	0.2106168	1.043	0.297045	
## factor(street)153	0.0696100	0.2107055	0.330	0.741170	
## factor(street)154	0.0355805	0.1413054	0.252	0.801232	
## factor(street)155	-0.1484092	0.2106588	-0.705	0.481232	
## factor(street)156	0.2905525	0.2106138	1.380	0.167933	
## factor(street)157	0.2408824	0.1890120	1.274	0.202712	
## factor(street)158	0.4923750	0.1744992	2.822	0.004842	**
## factor(street)160	-0.0268491	0.1378867	-0.195	0.845640	
## factor(street)163	0.0157202	0.1564268	0.100	0.919964	
## factor(street)164	-0.0027601	0.1744242	-0.016	0.987377	
## factor(street)165	0.1023135	0.1889925	0.541	0.588339	
## factor(street)166	-0.0511870	0.2106588	-0.243	0.808051	
## factor(street)168	0.2437562	0.2109999	1.155	0.248177	

## factor(street)170	0.0893236	0.1284410	0.695	0.486886	
## factor(street)171	0.0488373	0.1641402	0.298	0.766101	
## factor(street)172	0.0707283	0.1503022	0.471	0.638014	
## factor(street)175	0.3459145	0.1324845	2.611	0.009119	**
## factor(street)179	0.1009483	0.1889616	0.534	0.593265	
## factor(street)180	-0.0008258	0.1204432	-0.007	0.994531	
## factor(street)182	-0.0505141	0.1641171	-0.308	0.758283	
## factor(street)183	0.0124584	0.1214484	0.103	0.918309	
## factor(street)185	0.0329075	0.1641126	0.201	0.841103	
## factor(street)186	0.1008820	0.1503946	0.671	0.502464	
## factor(street)187	0.0840212	0.1302907	0.645	0.519108	
## factor(street)188	0.0673558	0.1453733	0.463	0.643196	
## factor(street)189	0.0402341	0.1203745	0.334	0.738245	
## factor(street)190	0.1053748	0.1504876	0.700	0.483899	
## factor(street)191	0.0798253	0.1453437	0.549	0.582939	
## factor(street)192	-0.0096192	0.1163220	-0.083	0.934106	
## factor(street)193	0.1582192	0.1503141	1.053	0.292700	
## factor(street)196	-0.1303779	0.1642063	-0.794	0.427329	
## factor(street)197	0.1682332	0.2106881	0.798	0.424712	
## factor(street)198	0.1803945	0.1284477	1.404	0.160404	
## factor(street)200	0.0507821	0.1564434	0.325	0.745527	
## factor(street)202	0.1053217	0.1746071	0.603	0.546473	
## factor(street)203	-0.1303425	0.1563558	-0.834	0.404626	
## factor(street)206	-0.1415574	0.1413584	-1.001	0.316793	
## factor(street)207	-0.1427886	0.1888010	-0.756	0.449595	
## factor(street)208	-0.0207241	0.1413584	-0.147	0.883463	
## factor(street)209	-0.2133688	0.1325220	-1.610	0.107597	
## factor(street)210	-0.1312847	0.1415290	-0.928	0.353758	
## factor(street)213	0.1372066	0.1350977	1.016	0.309981	
## factor(street)215	0.0373590	0.1135376	0.329	0.742168	
## factor(street)216	-0.3515156	0.1759400	-1.998	0.045907	*
## factor(street)217	-0.0845579	0.1502700	-0.563	0.573720	
## factor(street)220	-0.1100836	0.1563432	-0.704	0.481472	
## factor(street)221	-0.0544156	0.1453881	-0.374	0.708251	
## factor(street)222	-0.0803151	0.1744136	-0.460	0.645235	
## factor(street)223	-0.0829403	0.1503345	-0.552	0.581234	
## factor(street)225	0.0050526	0.1455381	0.035	0.972310	
## factor(street)227	0.0496890	0.1744684	0.285	0.775836	
## factor(street)228	0.1433788	0.2106199	0.681	0.496138	
## factor(street)229	0.1370082	0.1648917	0.831	0.406165	
## factor(street)230	0.1060271	0.1564049	0.678	0.497940	
## factor(street)232	0.1730376	0.1890159	0.915	0.360096	
## factor(street)233	0.1468207	0.1349779	1.088	0.276888	
## factor(street)235	-0.0192812	0.1745090	-0.110	0.912037	
## factor(street)236	0.0814638	0.1237519	0.658	0.510459	
## factor(street)238	-0.0144368	0.1453707	-0.099	0.920906	
## factor(street)240	-0.2020254	0.1744074	-1.158	0.246907	
## factor(street)241	-0.0093410	0.1888380	-0.049	0.960555	
## factor(street)242	0.1349727	0.2106216	0.641	0.521732	
## factor(street)243	0.2522588	0.1889839	1.335	0.182142	
## factor(street)244	0.1851309	0.1302822	1.421	0.155528	
## factor(street)246	-0.0426043	0.1642633	-0.259	0.795389	
## factor(street)247	0.0930582	0.1325007	0.702	0.482588	
## factor(street)248	0.0899696	0.1170122	0.769	0.442081	

```
## factor(street)249      0.1828708  0.1266888  1.443 0.149101
## factor(street)250      0.1412563  0.1118391  1.263 0.206777
## factor(street)253      0.2040057  0.1251248  1.630 0.103226
## factor(street)254      0.1977383  0.1565323  1.263 0.206701
## factor(street)255      0.1225908  0.1284542  0.954 0.340061
## factor(street)256      0.1227641  0.1132384  1.084 0.278488
## factor(street)257     -0.0063413  0.1238865  -0.051 0.959184
## factor(street)258      0.1905219  0.1203751  1.583 0.113695
## factor(street)259      0.7976129  0.2107826  3.784 0.000160 ***
## factor(street)260      0.0640570  0.1126272  0.569 0.569611
## factor(street)261      0.0511237  0.1503856  0.340 0.733942
## factor(street)262      0.1308689  0.1453818  0.900 0.368174
## factor(street)263      0.1233063  0.1251872  0.985 0.324797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3233 on 1480 degrees of freedom
## Multiple R-squared:  0.4421, Adjusted R-squared:  0.3743
## F-statistic: 6.516 on 180 and 1480 DF,  p-value: < 2.2e-16

paste("The treatment effect estimate is", signif(summary(model4c)$coefficients[2,1],3))

## [1] "The treatment effect estimate is 0.116"

paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0196"

paste("The 95% robust confidence interval is (", ans[1], ",", ans[2], ")")

## [1] "The 95% robust confidence interval is ( 0.0776 , 0.155 )"
```

- 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.

Adam Yang: Blocking by street would reduce the probability that large treatment and control group differences occur by chance by balancing the presence of similar units across the treatment and control groups. In other words, blocking reduces the sampling variability if the variability between blocks is greater than the variability within each block. The streets that each household lives on can represent the wealth of that household so we can expect there to be variability between blocks. In terms of the confidence interval, we performed random assignment based on the blocks so the impact from within these blocks are already intrinsically included in the model. Adding the covariate would therefore not provide much more new information so the confidence interval does not change much. In other words, adding a covariate would absorb variability in the model. However, when you’re blocking you are already minimizing that variability so adding the covariate would not have a large impact.

- 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 data set.

```
# Create new column with nocell variable
cleaned_d$nocell <- 1-cleaned_d$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.

```

model4f <- lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street) + nocell, data = cleaned_d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model4f, row = 2, z_val = 1.96)

summary(model4f)

```

```

##
## Call:
## lm(formula = avg_bins_treat ~ bin + base_avg_bins_treat + factor(street) +
##     nocell, data = cleaned_d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44271 -0.19428 -0.01711  0.16030  1.84323
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2876250  0.0978741   2.939 0.003347 **
## bin            0.1171694  0.0175861   6.663 3.78e-11 ***
## base_avg_bins_treat 0.3668400  0.0147461  24.877 < 2e-16 ***
## factor(street)3    0.0473170  0.1410830   0.335 0.737383
## factor(street)4   -0.0292088  0.1451975  -0.201 0.840596
## factor(street)5   -0.0009820  0.1451020  -0.007 0.994601
## factor(street)6    0.0430615  0.1500623   0.287 0.774185
## factor(street)7    0.0272138  0.1741994   0.156 0.875879
## factor(street)8    0.0444826  0.1300751   0.342 0.732417
## factor(street)9    0.0108438  0.1347546   0.080 0.935874
## factor(street)10   0.0697463  0.1223235   0.570 0.568643
## factor(street)11   0.0262337  0.1349941   0.194 0.845943
## factor(street)15   0.0726314  0.1886699   0.385 0.700318
## factor(street)17   0.1375254  0.1638485   0.839 0.401412
## factor(street)20  -0.1279096  0.1502520  -0.851 0.394740
## factor(street)21   0.0415966  0.1641703   0.253 0.800014
## factor(street)22  -0.0176789  0.1155863  -0.153 0.878459
## factor(street)23  -0.0540752  0.1176601  -0.460 0.645879
## factor(street)26   0.0865194  0.2104650   0.411 0.681068
## factor(street)32   0.2669652  0.2104820   1.268 0.204872
## factor(street)37   0.2553027  0.1742758   1.465 0.143151
## factor(street)38   0.0190313  0.1741396   0.109 0.912989
## factor(street)40   0.0359328  0.2104304   0.171 0.864437
## factor(street)41   0.1228156  0.1742599   0.705 0.481056
## factor(street)42  -0.0227996  0.2108171  -0.108 0.913892
## factor(street)43   0.1448257  0.1741330   0.832 0.405715
## factor(street)44   0.2172290  0.1501302   1.447 0.148126
## factor(street)45   0.0541653  0.1884842   0.287 0.773867
## factor(street)46   0.1026413  0.1639106   0.626 0.531278
## factor(street)47   0.5174350  0.1888403   2.740 0.006217 **
## factor(street)53   0.1204108  0.2103376   0.572 0.567095
## factor(street)58   0.0693009  0.1322832   0.524 0.600438
## factor(street)60   0.1269351  0.1264494   1.004 0.315619
## factor(street)61   0.0871902  0.2106536   0.414 0.679005
## factor(street)62   0.1406195  0.1561821   0.900 0.368077
## factor(street)63   0.0699528  0.1500131   0.466 0.641061

```


## factor(street)64	0.1421589	0.1638695	0.868	0.385802
## factor(street)66	0.0486430	0.1561965	0.311	0.755524
## factor(street)67	0.0957189	0.1452293	0.659	0.509942
## factor(street)68	0.1832636	0.1501320	1.221	0.222401
## factor(street)69	0.1686643	0.1377487	1.224	0.220983
## factor(street)70	0.0804385	0.1300919	0.618	0.536459
## factor(street)72	0.0581678	0.1378181	0.422	0.673041
## factor(street)73	-0.0772537	0.1500655	-0.515	0.606770
## factor(street)74	0.0280909	0.1211786	0.232	0.816715
## factor(street)75	0.0511721	0.1348826	0.379	0.704458
## factor(street)77	0.0611698	0.1561638	0.392	0.695334
## factor(street)78	0.0416203	0.1264378	0.329	0.742069
## factor(street)79	0.4223495	0.1887215	2.238	0.025373 *
## factor(street)80	0.1127090	0.1155049	0.976	0.329326
## factor(street)81	0.0710531	0.1885893	0.377	0.706405
## factor(street)82	0.1290389	0.2103730	0.613	0.539718
## factor(street)83	-0.0203079	0.1564437	-0.130	0.896735
## factor(street)84	0.0958025	0.1562674	0.613	0.539926
## factor(street)85	0.1666365	0.1264694	1.318	0.187841
## factor(street)86	0.0290504	0.1224034	0.237	0.812431
## factor(street)88	0.0003815	0.1411286	0.003	0.997844
## factor(street)89	0.0884200	0.1500615	0.589	0.555800
## factor(street)91	-0.1295170	0.2103351	-0.616	0.538144
## factor(street)93	0.0312294	0.1561181	0.200	0.841479
## factor(street)94	0.0492609	0.1451620	0.339	0.734393
## factor(street)96	0.0621879	0.1561215	0.398	0.690444
## factor(street)98	0.1423894	0.1348878	1.056	0.291317
## factor(street)99	0.3333284	0.1453050	2.294	0.021931 *
## factor(street)100	0.1740860	0.1413143	1.232	0.218180
## factor(street)101	0.3110685	0.1453525	2.140	0.032510 *
## factor(street)102	0.0506541	0.2118178	0.239	0.811030
## factor(street)103	0.1872478	0.1451332	1.290	0.197190
## factor(street)105	0.2183671	0.1742314	1.253	0.210289
## factor(street)106	0.4916118	0.1348038	3.647	0.000275 ***
## factor(street)107	0.1485830	0.1885585	0.788	0.430826
## factor(street)109	0.0807446	0.1500390	0.538	0.590549
## factor(street)110	0.2245482	0.1742125	1.289	0.197623
## factor(street)111	0.0861889	0.1742618	0.495	0.620960
## factor(street)112	0.2885674	0.1503465	1.919	0.055132 .
## factor(street)113	0.1654387	0.1564169	1.058	0.290375
## factor(street)115	0.3816887	0.1349054	2.829	0.004728 **
## factor(street)117	0.6807328	0.2119636	3.212	0.001349 **
## factor(street)118	0.1176185	0.2103597	0.559	0.576157
## factor(street)119	0.0842458	0.1500621	0.561	0.574606
## factor(street)120	0.0232633	0.1501161	0.155	0.876867
## factor(street)121	0.2759211	0.1502491	1.836	0.066496 .
## factor(street)122	-0.0646606	0.1885460	-0.343	0.731690
## factor(street)124	0.1266702	0.1376998	0.920	0.357774
## factor(street)125	0.4014512	0.1270225	3.160	0.001607 **
## factor(street)126	-0.0158853	0.1643710	-0.097	0.923023
## factor(street)127	0.1762915	0.1643050	1.073	0.283467
## factor(street)128	0.2512148	0.1348946	1.862	0.062758 .
## factor(street)129	0.4761889	0.2104819	2.262	0.023819 *
## factor(street)130	0.2776131	0.1501427	1.849	0.064658 .

## factor(street)131	0.3741923	0.1505384	2.486	0.013040	*
## factor(street)132	0.3357778	0.1284528	2.614	0.009039	**
## factor(street)133	0.4209772	0.1885869	2.232	0.025746	*
## factor(street)134	0.0934959	0.1562717	0.598	0.549737	
## factor(street)136	0.2682935	0.1503800	1.784	0.074612	.
## factor(street)137	0.3541631	0.2103075	1.684	0.092388	.
## factor(street)138	0.2119487	0.1382845	1.533	0.125564	
## factor(street)147	0.2901211	0.2103950	1.379	0.168123	
## factor(street)148	0.2009915	0.1884986	1.066	0.286473	
## factor(street)149	0.3774627	0.1454622	2.595	0.009555	**
## factor(street)151	-0.1611941	0.2105575	-0.766	0.444061	
## factor(street)152	0.2365772	0.2103760	1.125	0.260964	
## factor(street)153	0.0865194	0.2104650	0.411	0.681068	
## factor(street)154	0.0452158	0.1411229	0.320	0.748710	
## factor(street)155	-0.1601593	0.2103597	-0.761	0.446564	
## factor(street)156	0.3074178	0.2103729	1.461	0.144146	
## factor(street)157	0.2500940	0.1887320	1.325	0.185334	
## factor(street)158	0.4888961	0.1742118	2.806	0.005077	**
## factor(street)160	-0.0191333	0.1376912	-0.139	0.889502	
## factor(street)163	0.0036677	0.1562416	0.023	0.981275	
## factor(street)164	0.0111749	0.1742242	0.064	0.948867	
## factor(street)165	0.1115361	0.1887126	0.591	0.554586	
## factor(street)166	-0.0342922	0.2104182	-0.163	0.870563	
## factor(street)168	0.2598748	0.2107484	1.233	0.217733	
## factor(street)170	0.0977005	0.1282709	0.762	0.446376	
## factor(street)171	0.0657222	0.1640097	0.401	0.688683	
## factor(street)172	0.0911515	0.1502817	0.607	0.544251	
## factor(street)175	0.3607661	0.1324011	2.725	0.006510	**
## factor(street)179	0.0890336	0.1887069	0.472	0.637133	
## factor(street)180	0.0079402	0.1202942	0.066	0.947382	
## factor(street)182	-0.0477894	0.1638452	-0.292	0.770576	
## factor(street)183	0.0284945	0.1214214	0.235	0.814494	
## factor(street)185	0.0426141	0.1638849	0.260	0.794881	
## factor(street)186	0.0998437	0.1501425	0.665	0.506159	
## factor(street)187	0.0998979	0.1302335	0.767	0.443164	
## factor(street)188	0.0699183	0.1451328	0.482	0.630052	
## factor(street)189	0.0491278	0.1202272	0.409	0.682874	
## factor(street)190	0.1199941	0.1503536	0.798	0.424952	
## factor(street)191	0.0824959	0.1451036	0.569	0.569761	
## factor(street)192	-0.0015741	0.1161731	-0.014	0.989191	
## factor(street)193	0.1786460	0.1502936	1.189	0.234769	
## factor(street)196	-0.1349570	0.1639411	-0.823	0.410525	
## factor(street)197	0.1564930	0.2103889	0.744	0.457099	
## factor(street)198	0.1917096	0.1283153	1.494	0.135376	
## factor(street)200	0.0633054	0.1562644	0.405	0.685450	
## factor(street)202	0.1281242	0.1745627	0.734	0.463082	
## factor(street)203	-0.1359179	0.1561098	-0.871	0.384083	
## factor(street)206	-0.1361967	0.1411379	-0.965	0.334707	
## factor(street)207	-0.1330096	0.1885262	-0.706	0.480596	
## factor(street)208	-0.0024732	0.1413179	-0.018	0.986039	
## factor(street)209	-0.2118752	0.1323008	-1.601	0.109487	
## factor(street)210	-0.1085927	0.1415953	-0.767	0.443250	
## factor(street)213	0.1538054	0.1350413	1.139	0.254908	
## factor(street)215	0.0478623	0.1134281	0.422	0.673115	

```
## factor(street)216 -0.3460032 0.1756588 -1.970 0.049054 *
## factor(street)217 -0.0853715 0.1500179 -0.569 0.569392
## factor(street)220 -0.0972529 0.1561686 -0.623 0.533550
## factor(street)221 -0.0566223 0.1451466 -0.390 0.696516
## factor(street)222 -0.0575921 0.1743681 -0.330 0.741227
## factor(street)223 -0.0732497 0.1501342 -0.488 0.625696
## factor(street)225 0.0120225 0.1453215 0.083 0.934077
## factor(street)227 0.0464577 0.1741802 0.267 0.789721
## factor(street)228 0.1459020 0.2102685 0.694 0.487864
## factor(street)229 0.1536155 0.1647545 0.932 0.351287
## factor(street)230 0.1125807 0.1561651 0.721 0.471080
## factor(street)232 0.1823045 0.1887363 0.966 0.334241
## factor(street)233 0.1493562 0.1347551 1.108 0.267890
## factor(street)235 -0.0225920 0.1742210 -0.130 0.896842
## factor(street)236 0.0984495 0.1237389 0.796 0.426379
## factor(street)238 -0.0024832 0.1452086 -0.017 0.986359
## factor(street)240 -0.2050651 0.1741188 -1.178 0.239093
## factor(street)241 -0.0105654 0.1885214 -0.056 0.955315
## factor(street)242 0.1518478 0.2103808 0.722 0.470546
## factor(street)243 0.2400200 0.1887327 1.272 0.203663
## factor(street)244 0.2040852 0.1302938 1.566 0.117481
## factor(street)246 -0.0260976 0.1641261 -0.159 0.873683
## factor(street)247 0.0978503 0.1322925 0.740 0.459629
## factor(street)248 0.1006470 0.1168971 0.861 0.389384
## factor(street)249 0.1954384 0.1265802 1.544 0.122805
## factor(street)250 0.1527487 0.1117499 1.367 0.171871
## factor(street)253 0.2200733 0.1250871 1.759 0.078723 .
## factor(street)254 0.2102890 0.1563535 1.345 0.178845
## factor(street)255 0.1363813 0.1283622 1.062 0.288195
## factor(street)256 0.1367071 0.1131917 1.208 0.227337
## factor(street)257 0.0010780 0.1237155 0.009 0.993049
## factor(street)258 0.1952463 0.1201883 1.625 0.104482
## factor(street)259 0.8284665 0.2108060 3.930 8.89e-05 ***
## factor(street)260 0.0667213 0.1124432 0.593 0.553019
## factor(street)261 0.0658849 0.1502541 0.438 0.661095
## factor(street)262 0.1476228 0.1452989 1.016 0.309798
## factor(street)263 0.1318166 0.1250252 1.054 0.291909
## nocell -0.0429672 0.0175619 -2.447 0.014536 *
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 0.3228 on 1479 degrees of freedom
```

```
## Multiple R-squared: 0.4444, Adjusted R-squared: 0.3764
```

```
## F-statistic: 6.535 on 181 and 1479 DF, p-value: < 2.2e-16
```

```
paste("The treatment effect estimate is", signif(summary(model4f)$coefficients[2,1],3))
```

```
## [1] "The treatment effect estimate is 0.117"
```

```
paste("The robust standard error is", ans[3])
```

```
## [1] "The robust standard error is 0.0196"
```

```
paste("The 95% robust confidence interval is (", ans[1],",",ans[2],",)")
```

```
## [1] "The 95% robust confidence interval is ( 0.0787 , 0.156 )"
```

Adam Yang: I think it is very likely that having a cell phone or not does not help explain the variability in the treatment effect of having a recycling bin. Therefore, because having a cell phone should not have much to do with having a recycling bin and the recycling intensity, it does not help tighten our confidence interval. It's also possible that having a cell phone or not also represents the economic class of the household. This information is already provided in some way by the `street` variable, so not much more information is provided.

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.

Note to Grader: I used the cleaned up data set which removed the data points with `street = -999` so the results are not exactly like the results in Table 4A. I did use the original data set to confirm that I got the same results as Table 4A if I did not clean the data.

```
model4g <- lm(avg_bins_treat ~ bin + sms + base_avg_bins_treat + factor(street) + nocell, data = cleaned_d)
```

```
# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model4g, row = 2, z_val = 1.96)
```

```
summary(model4g)
```

```
##
## Call:
## lm(formula = avg_bins_treat ~ bin + sms + base_avg_bins_treat +
##      factor(street) + nocell, data = cleaned_d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43649 -0.19442 -0.01749  0.16096  1.85315
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.278621   0.098526   2.828 0.004749 **
## bin            0.116968   0.017590   6.650 4.12e-11 ***
## sms            0.017400   0.021666   0.803 0.422039
## base_avg_bins_treat 0.367082   0.014751  24.885 < 2e-16 ***
## factor(street)3    0.045630   0.141116   0.323 0.746475
## factor(street)4   -0.031352   0.145239  -0.216 0.829123
## factor(street)5   -0.001938   0.145124  -0.013 0.989346
## factor(street)6    0.040918   0.150104   0.273 0.785200
## factor(street)7    0.025733   0.174230   0.148 0.882603
## factor(street)8    0.043776   0.130094   0.336 0.736546
## factor(street)9    0.010774   0.134771   0.080 0.936295
## factor(street)10   0.067395   0.122373   0.551 0.581902
## factor(street)11   0.023957   0.135040   0.177 0.859212
## factor(street)15   0.074900   0.188714   0.397 0.691501
## factor(street)17   0.137455   0.163868   0.839 0.401711
## factor(street)20  -0.130408   0.150302  -0.868 0.385733
## factor(street)21   0.039734   0.164206   0.242 0.808834
## factor(street)22  -0.018595   0.115606  -0.161 0.872237
## factor(street)23  -0.055055   0.117681  -0.468 0.639975
## factor(street)26   0.083197   0.210531   0.395 0.692767
## factor(street)32   0.263791   0.210544   1.253 0.210440
## factor(street)37   0.253842   0.174306   1.456 0.145522
## factor(street)38   0.015671   0.174211   0.090 0.928338
```

## factor(street)40	0.035879	0.210456	0.170	0.864655	
## factor(street)41	0.122501	0.174281	0.703	0.482232	
## factor(street)42	-0.023244	0.210843	-0.110	0.912233	
## factor(street)43	0.144891	0.174154	0.832	0.405561	
## factor(street)44	0.215315	0.150167	1.434	0.151831	
## factor(street)45	0.052028	0.188526	0.276	0.782607	
## factor(street)46	0.101187	0.163940	0.617	0.537186	
## factor(street)47	0.519494	0.188880	2.750	0.006025	**
## factor(street)53	0.117700	0.210390	0.559	0.575947	
## factor(street)58	0.066526	0.132344	0.503	0.615270	
## factor(street)60	0.125768	0.126473	0.994	0.320177	
## factor(street)61	0.087254	0.210679	0.414	0.678821	
## factor(street)62	0.142097	0.156212	0.910	0.363158	
## factor(street)63	0.066623	0.150088	0.444	0.657186	
## factor(street)64	0.143663	0.163900	0.877	0.380884	
## factor(street)66	0.047088	0.156227	0.301	0.763148	
## factor(street)67	0.095408	0.145247	0.657	0.511367	
## factor(street)68	0.181849	0.150160	1.211	0.226076	
## factor(street)69	0.167978	0.137768	1.219	0.222933	
## factor(street)70	0.078072	0.130141	0.600	0.548661	
## factor(street)72	0.056319	0.137854	0.409	0.682938	
## factor(street)73	-0.079401	0.150107	-0.529	0.596912	
## factor(street)74	0.027693	0.121194	0.229	0.819287	
## factor(street)75	0.048743	0.134933	0.361	0.717975	
## factor(street)77	0.059967	0.156190	0.384	0.701080	
## factor(street)78	0.040564	0.126460	0.321	0.748431	
## factor(street)79	0.424124	0.188757	2.247	0.024792	*
## factor(street)80	0.112564	0.115519	0.974	0.330008	
## factor(street)81	0.066611	0.188693	0.353	0.724129	
## factor(street)82	0.134855	0.210523	0.641	0.521899	
## factor(street)83	-0.022777	0.156493	-0.146	0.884298	
## factor(street)84	0.095429	0.156287	0.611	0.541559	
## factor(street)85	0.165956	0.126488	1.312	0.189712	
## factor(street)86	0.027169	0.122440	0.222	0.824429	
## factor(street)88	-0.002325	0.141186	-0.016	0.986862	
## factor(street)89	0.086341	0.150102	0.575	0.565233	
## factor(street)91	-0.126327	0.210398	-0.600	0.548320	
## factor(street)93	0.031085	0.156137	0.199	0.842224	
## factor(street)94	0.045593	0.145251	0.314	0.753647	
## factor(street)96	0.062052	0.156140	0.397	0.691122	
## factor(street)98	0.146857	0.135019	1.088	0.276914	
## factor(street)99	0.334188	0.145326	2.300	0.021612	*
## factor(street)100	0.173056	0.141337	1.224	0.220990	
## factor(street)101	0.313563	0.145403	2.157	0.031205	*
## factor(street)102	0.049690	0.211847	0.235	0.814586	
## factor(street)103	0.186174	0.145157	1.283	0.199844	
## factor(street)105	0.216451	0.174269	1.242	0.214414	
## factor(street)106	0.494546	0.134870	3.667	0.000254	***
## factor(street)107	0.153114	0.188666	0.812	0.417175	
## factor(street)109	0.085043	0.150152	0.566	0.571222	
## factor(street)110	0.229977	0.174365	1.319	0.187391	
## factor(street)111	0.088102	0.174299	0.505	0.613309	
## factor(street)112	0.291574	0.150411	1.939	0.052751	.
## factor(street)113	0.169408	0.156514	1.082	0.279259	

## factor(street)115	0.382464	0.134925	2.835	0.004650	**
## factor(street)117	0.682889	0.212006	3.221	0.001305	**
## factor(street)118	0.120688	0.210420	0.574	0.566353	
## factor(street)119	0.088568	0.150177	0.590	0.555441	
## factor(street)120	0.028717	0.150288	0.191	0.848490	
## factor(street)121	0.280143	0.150359	1.863	0.062638	.
## factor(street)122	-0.062553	0.188587	-0.332	0.740167	
## factor(street)124	0.131563	0.137851	0.954	0.340045	
## factor(street)125	0.402858	0.127050	3.171	0.001551	**
## factor(street)126	-0.010003	0.164554	-0.061	0.951536	
## factor(street)127	0.178950	0.164358	1.089	0.276429	
## factor(street)128	0.252433	0.134919	1.871	0.061544	.
## factor(street)129	0.476105	0.210507	2.262	0.023861	*
## factor(street)130	0.280709	0.150210	1.869	0.061852	.
## factor(street)131	0.378408	0.150648	2.512	0.012115	*
## factor(street)132	0.339074	0.128534	2.638	0.008427	**
## factor(street)133	0.425130	0.188680	2.253	0.024394	*
## factor(street)134	0.093284	0.156291	0.597	0.550693	
## factor(street)136	0.268361	0.150398	1.784	0.074574	.
## factor(street)137	0.354052	0.210333	1.683	0.092529	.
## factor(street)138	0.216737	0.138430	1.566	0.117637	
## factor(street)147	0.295897	0.210543	1.405	0.160112	
## factor(street)148	0.207486	0.188695	1.100	0.271690	
## factor(street)149	0.382246	0.145601	2.625	0.008747	**
## factor(street)151	-0.161416	0.210583	-0.767	0.443490	
## factor(street)152	0.239136	0.210425	1.136	0.255957	
## factor(street)153	0.083197	0.210531	0.395	0.692767	
## factor(street)154	0.044158	0.141146	0.313	0.754435	
## factor(street)155	-0.162890	0.210412	-0.774	0.438969	
## factor(street)156	0.304186	0.210437	1.446	0.148529	
## factor(street)157	0.250053	0.188755	1.325	0.185459	
## factor(street)158	0.489007	0.174233	2.807	0.005072	**
## factor(street)160	-0.019276	0.137708	-0.140	0.888695	
## factor(street)163	0.005209	0.156272	0.033	0.973416	
## factor(street)164	0.007417	0.174308	0.043	0.966065	
## factor(street)165	0.111472	0.188735	0.591	0.554862	
## factor(street)166	-0.037584	0.210483	-0.179	0.858307	
## factor(street)168	0.262598	0.210801	1.246	0.213066	
## factor(street)170	0.095140	0.128326	0.741	0.458574	
## factor(street)171	0.065351	0.164030	0.398	0.690388	
## factor(street)172	0.090711	0.150301	0.604	0.546251	
## factor(street)175	0.359790	0.132423	2.717	0.006665	**
## factor(street)179	0.084846	0.188802	0.449	0.653213	
## factor(street)180	0.007279	0.120312	0.061	0.951763	
## factor(street)182	-0.050769	0.163907	-0.310	0.756801	
## factor(street)183	0.026863	0.121453	0.221	0.824982	
## factor(street)185	0.040968	0.163917	0.250	0.802675	
## factor(street)186	0.099778	0.150161	0.664	0.506488	
## factor(street)187	0.098993	0.130254	0.760	0.447376	
## factor(street)188	0.069827	0.145150	0.481	0.630538	
## factor(street)189	0.048500	0.120244	0.403	0.686753	
## factor(street)190	0.118684	0.150381	0.789	0.430106	
## factor(street)191	0.082413	0.145121	0.568	0.570196	
## factor(street)192	-0.002415	0.116192	-0.021	0.983419	

```

## factor(street)193    0.178198    0.150313    1.186 0.236005
## factor(street)196   -0.133498    0.163971   -0.814 0.415687
## factor(street)197    0.153742    0.210442    0.731 0.465158
## factor(street)198    0.190819    0.128335    1.487 0.137260
## factor(street)200    0.064318    0.156288    0.412 0.680741
## factor(street)202    0.125789    0.174608    0.720 0.471388
## factor(street)203   -0.133359    0.156161   -0.854 0.393251
## factor(street)206   -0.138085    0.141174   -0.978 0.328175
## factor(street)207   -0.133182    0.188549   -0.706 0.480080
## factor(street)208   -0.003744    0.141344   -0.026 0.978868
## factor(street)209   -0.213967    0.132342   -1.617 0.106142
## factor(street)210   -0.109168    0.141614   -0.771 0.440900
## factor(street)213    0.152054    0.135075    1.126 0.260474
## factor(street)215    0.046192    0.113461    0.407 0.683979
## factor(street)216   -0.344723    0.175687   -1.962 0.049933 *
## factor(street)217   -0.087556    0.150061   -0.583 0.559663
## factor(street)220   -0.098766    0.156199   -0.632 0.527280
## factor(street)221   -0.059530    0.145209   -0.410 0.681895
## factor(street)222   -0.056284    0.174397   -0.323 0.746940
## factor(street)223   -0.073499    0.150152   -0.489 0.624563
## factor(street)225    0.009002    0.145388    0.062 0.950638
## factor(street)227    0.046474    0.174201    0.267 0.789672
## factor(street)228    0.140092    0.210418    0.666 0.505658
## factor(street)229    0.153125    0.164775    0.929 0.352888
## factor(street)230    0.112420    0.156184    0.720 0.471766
## factor(street)232    0.177800    0.188842    0.942 0.346590
## factor(street)233    0.147871    0.134784    1.097 0.272779
## factor(street)235   -0.022412    0.174242   -0.129 0.897670
## factor(street)236    0.098101    0.123755    0.793 0.428077
## factor(street)238   -0.002668    0.145226   -0.018 0.985343
## factor(street)240   -0.208509    0.174192   -1.197 0.231498
## factor(street)241   -0.012682    0.188562   -0.067 0.946385
## factor(street)242    0.148596    0.210445    0.706 0.480234
## factor(street)243    0.244681    0.188845    1.296 0.195289
## factor(street)244    0.203724    0.130310    1.563 0.118178
## factor(street)246   -0.026382    0.164146   -0.161 0.872335
## factor(street)247    0.096377    0.132321    0.728 0.466510
## factor(street)248    0.100398    0.116912    0.859 0.390616
## factor(street)249    0.193499    0.126618    1.528 0.126674
## factor(street)250    0.151499    0.111774    1.355 0.175497
## factor(street)253    0.219748    0.125103    1.757 0.079204 .
## factor(street)254    0.211245    0.156377    1.351 0.176943
## factor(street)255    0.136124    0.128378    1.060 0.289163
## factor(street)256    0.136034    0.113208    1.202 0.229700
## factor(street)257   -0.000128    0.123740   -0.001 0.999175
## factor(street)258    0.195587    0.120204    1.627 0.103922
## factor(street)259    0.827851    0.210833    3.927 9.01e-05 ***
## factor(street)260    0.066609    0.112457    0.592 0.553734
## factor(street)261    0.064542    0.150281    0.429 0.667639
## factor(street)262    0.148257    0.145318    1.020 0.307790
## factor(street)263    0.131076    0.125044    1.048 0.294698
## nocell              -0.033425    0.021205   -1.576 0.115174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##
## Residual standard error: 0.3228 on 1478 degrees of freedom
## Multiple R-squared:  0.4446, Adjusted R-squared:  0.3762
## F-statistic: 6.501 on 182 and 1478 DF,  p-value: < 2.2e-16
paste("The treatment effect estimate is", signif(summary(model4g)$coefficients[2,1],3))

## [1] "The treatment effect estimate is 0.117"
paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0196"
paste("The 95% robust confidence interval is (", ans[1], ",", ans[2], ")")

## [1] "The 95% robust confidence interval is ( 0.0785 , 0.155 )"
```

Adam Yang: The confidence interval did not seem to change too much. The confidence interval we are showing is the confidence interval around the estimated treatment effect of having a recycling bin or not. It seems like receiving an SMS text message does not decrease the variance of the treatment effect of having a recycling bin or not. Also, we already included the covariate of `nocell` which should be correlated with the `sms` covariate so we may not be getting much new information and thus the confidence interval does not change by much.

- 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.

```
model4h <- lm(avg_bins_treat ~ bin_g + bin_s + sms_p + sms_g + nocell + base_avg_bins_treat
              + factor(street), data = cleaned_d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model4h, row = 2, z_val = 1.96)

summary(model4h)

##
## Call:
## lm(formula = avg_bins_treat ~ bin_g + bin_s + sms_p + sms_g +
##     nocell + base_avg_bins_treat + factor(street), data = cleaned_d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44451 -0.19769 -0.01497  0.16196  1.85274
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.279069   0.098554   2.832 0.004694 **
## bin_g          0.105739   0.022723   4.653 3.56e-06 ***
## bin_s          0.128862   0.022730   5.669 1.72e-08 ***
## sms_p          0.006660   0.026030    0.256 0.798107
## sms_g          0.029504   0.026162    1.128 0.259608
## nocell        -0.033157   0.021212  -1.563 0.118229
## base_avg_bins_treat 0.367440   0.014758  24.898 < 2e-16 ***
## factor(street)3    0.045555   0.141164    0.323 0.746962
## factor(street)4   -0.033654   0.145289  -0.232 0.816857
## factor(street)5   -0.003047   0.145166  -0.021 0.983258
```


## factor(street)6	0.042546	0.150156	0.283	0.776954
## factor(street)7	0.024656	0.174350	0.141	0.887559
## factor(street)8	0.042732	0.130130	0.328	0.742671
## factor(street)9	0.008750	0.134827	0.065	0.948265
## factor(street)10	0.068009	0.122405	0.556	0.578562
## factor(street)11	0.020676	0.135131	0.153	0.878416
## factor(street)15	0.067995	0.188860	0.360	0.718877
## factor(street)17	0.132486	0.164023	0.808	0.419375
## factor(street)20	-0.130282	0.150356	-0.866	0.386364
## factor(street)21	0.030926	0.164483	0.188	0.850886
## factor(street)22	-0.018388	0.115636	-0.159	0.873675
## factor(street)23	-0.058599	0.117771	-0.498	0.618861
## factor(street)26	0.089551	0.210660	0.425	0.670828
## factor(street)32	0.266392	0.210817	1.264	0.206568
## factor(street)37	0.247918	0.174662	1.419	0.155989
## factor(street)38	0.012199	0.174284	0.070	0.944207
## factor(street)40	0.046094	0.210716	0.219	0.826874
## factor(street)41	0.114428	0.174493	0.656	0.512073
## factor(street)42	-0.024241	0.210898	-0.115	0.908507
## factor(street)43	0.139135	0.174341	0.798	0.424963
## factor(street)44	0.212599	0.150299	1.415	0.157423
## factor(street)45	0.053823	0.188585	0.285	0.775375
## factor(street)46	0.103830	0.164001	0.633	0.526761
## factor(street)47	0.509547	0.189160	2.694	0.007146 **
## factor(street)53	0.112564	0.210536	0.535	0.592968
## factor(street)58	0.066191	0.132377	0.500	0.617138
## factor(street)60	0.125390	0.126505	0.991	0.321757
## factor(street)61	0.093579	0.210806	0.444	0.657172
## factor(street)62	0.144307	0.156340	0.923	0.356140
## factor(street)63	0.065494	0.150132	0.436	0.662725
## factor(street)64	0.142571	0.163945	0.870	0.384646
## factor(street)66	0.047653	0.156267	0.305	0.760452
## factor(street)67	0.091733	0.145353	0.631	0.528069
## factor(street)68	0.180751	0.150203	1.203	0.229024
## factor(street)69	0.167796	0.137814	1.218	0.223588
## factor(street)70	0.077702	0.130173	0.597	0.550659
## factor(street)72	0.055107	0.137895	0.400	0.689490
## factor(street)73	-0.080669	0.150151	-0.537	0.591173
## factor(street)74	0.026633	0.121231	0.220	0.826142
## factor(street)75	0.045779	0.135017	0.339	0.734610
## factor(street)77	0.060387	0.156246	0.386	0.699190
## factor(street)78	0.036581	0.126559	0.289	0.772590
## factor(street)79	0.417270	0.188990	2.208	0.027405 *
## factor(street)80	0.111103	0.115562	0.961	0.336502
## factor(street)81	0.065410	0.188744	0.347	0.728977
## factor(street)82	0.141595	0.210747	0.672	0.501769
## factor(street)83	-0.022433	0.156548	-0.143	0.886077
## factor(street)84	0.096070	0.156366	0.614	0.539049
## factor(street)85	0.162736	0.126617	1.285	0.198902
## factor(street)86	0.024689	0.122494	0.202	0.840297
## factor(street)88	0.001007	0.141255	0.007	0.994311
## factor(street)89	0.088064	0.150155	0.586	0.557640
## factor(street)91	-0.134894	0.210587	-0.641	0.521909
## factor(street)93	0.028455	0.156211	0.182	0.855484

## factor(street)94	0.046933	0.145297	0.323	0.746730	
## factor(street)96	0.062738	0.156181	0.402	0.687960	
## factor(street)98	0.144851	0.135073	1.072	0.283720	
## factor(street)99	0.332994	0.145368	2.291	0.022121	*
## factor(street)100	0.172908	0.141372	1.223	0.221497	
## factor(street)101	0.311223	0.145466	2.139	0.032560	*
## factor(street)102	0.051986	0.211923	0.245	0.806254	
## factor(street)103	0.185099	0.145230	1.275	0.202680	
## factor(street)105	0.215159	0.174318	1.234	0.217290	
## factor(street)106	0.493650	0.134907	3.659	0.000262	***
## factor(street)107	0.152120	0.188809	0.806	0.420557	
## factor(street)109	0.081276	0.150229	0.541	0.588576	
## factor(street)110	0.235909	0.174490	1.352	0.176585	
## factor(street)111	0.082310	0.174487	0.472	0.637192	
## factor(street)112	0.291883	0.150478	1.940	0.052605	.
## factor(street)113	0.163266	0.156729	1.042	0.297717	
## factor(street)115	0.382285	0.134971	2.832	0.004683	**
## factor(street)117	0.677477	0.212154	3.193	0.001436	**
## factor(street)118	0.119650	0.210558	0.568	0.569950	
## factor(street)119	0.084645	0.150369	0.563	0.573580	
## factor(street)120	0.027818	0.150336	0.185	0.853224	
## factor(street)121	0.277594	0.150485	1.845	0.065288	.
## factor(street)122	-0.063601	0.188683	-0.337	0.736107	
## factor(street)124	0.132721	0.137892	0.962	0.335959	
## factor(street)125	0.402287	0.127082	3.166	0.001579	**
## factor(street)126	-0.007378	0.164613	-0.045	0.964258	
## factor(street)127	0.175792	0.164424	1.069	0.285185	
## factor(street)128	0.248413	0.135041	1.840	0.066036	.
## factor(street)129	0.478568	0.210614	2.272	0.023214	*
## factor(street)130	0.279627	0.150260	1.861	0.062950	.
## factor(street)131	0.378582	0.150699	2.512	0.012105	*
## factor(street)132	0.340242	0.128575	2.646	0.008225	**
## factor(street)133	0.418332	0.188911	2.214	0.026951	*
## factor(street)134	0.093677	0.156330	0.599	0.549115	
## factor(street)136	0.267063	0.150442	1.775	0.076073	.
## factor(street)137	0.345217	0.210531	1.640	0.101272	
## factor(street)138	0.218618	0.138488	1.579	0.114641	
## factor(street)147	0.294870	0.210598	1.400	0.161677	
## factor(street)148	0.209521	0.188757	1.110	0.267179	
## factor(street)149	0.375911	0.145939	2.576	0.010097	*
## factor(street)151	-0.170415	0.210787	-0.808	0.418949	
## factor(street)152	0.242028	0.210508	1.150	0.250440	
## factor(street)153	0.081843	0.210729	0.388	0.697791	
## factor(street)154	0.041873	0.141195	0.297	0.766846	
## factor(street)155	-0.168056	0.210559	-0.798	0.424918	
## factor(street)156	0.310674	0.210569	1.475	0.140318	
## factor(street)157	0.237412	0.189218	1.255	0.209787	
## factor(street)158	0.490131	0.174282	2.812	0.004984	**
## factor(street)160	-0.019318	0.137754	-0.140	0.888492	
## factor(street)163	-0.004139	0.156613	-0.026	0.978918	
## factor(street)164	0.006170	0.174357	0.035	0.971777	
## factor(street)165	0.110290	0.188876	0.584	0.559358	
## factor(street)166	-0.038801	0.210621	-0.184	0.853865	
## factor(street)168	0.265314	0.210885	1.258	0.208555	

## factor(street)170	0.093901	0.128366	0.732	0.464584	
## factor(street)171	0.062281	0.164201	0.379	0.704523	
## factor(street)172	0.092480	0.150406	0.615	0.538737	
## factor(street)175	0.359665	0.132466	2.715	0.006702	**
## factor(street)179	0.080507	0.188894	0.426	0.670022	
## factor(street)180	0.008309	0.120349	0.069	0.944966	
## factor(street)182	-0.048073	0.163968	-0.293	0.769420	
## factor(street)183	0.025670	0.121487	0.211	0.832682	
## factor(street)185	0.035978	0.164073	0.219	0.826463	
## factor(street)186	0.098519	0.150242	0.656	0.512097	
## factor(street)187	0.097162	0.130296	0.746	0.455965	
## factor(street)188	0.071237	0.145194	0.491	0.623759	
## factor(street)189	0.045815	0.120321	0.381	0.703427	
## factor(street)190	0.117425	0.150423	0.781	0.435145	
## factor(street)191	0.080081	0.145171	0.552	0.581284	
## factor(street)192	-0.002210	0.116221	-0.019	0.984829	
## factor(street)193	0.177030	0.150392	1.177	0.239334	
## factor(street)196	-0.134657	0.164017	-0.821	0.411781	
## factor(street)197	0.148547	0.210589	0.705	0.480682	
## factor(street)198	0.190446	0.128368	1.484	0.138128	
## factor(street)200	0.064812	0.156343	0.415	0.678534	
## factor(street)202	0.124558	0.174683	0.713	0.475927	
## factor(street)203	-0.137688	0.156292	-0.881	0.378477	
## factor(street)206	-0.139337	0.141216	-0.987	0.323954	
## factor(street)207	-0.134209	0.188693	-0.711	0.477038	
## factor(street)208	-0.007253	0.141442	-0.051	0.959111	
## factor(street)209	-0.215197	0.132383	-1.626	0.104256	
## factor(street)210	-0.105781	0.141685	-0.747	0.455427	
## factor(street)213	0.148909	0.135139	1.102	0.270687	
## factor(street)215	0.045767	0.113490	0.403	0.686806	
## factor(street)216	-0.339391	0.175797	-1.931	0.053726	.
## factor(street)217	-0.090089	0.150131	-0.600	0.548551	
## factor(street)220	-0.099881	0.156243	-0.639	0.522746	
## factor(street)221	-0.063364	0.145287	-0.436	0.662805	
## factor(street)222	-0.054853	0.174448	-0.314	0.753233	
## factor(street)223	-0.076129	0.150222	-0.507	0.612388	
## factor(street)225	0.007636	0.145431	0.053	0.958130	
## factor(street)227	0.040647	0.174390	0.233	0.815732	
## factor(street)228	0.142693	0.210495	0.678	0.497945	
## factor(street)229	0.153522	0.164836	0.931	0.351817	
## factor(street)230	0.112890	0.156224	0.723	0.470029	
## factor(street)232	0.179238	0.188900	0.949	0.342852	
## factor(street)233	0.147704	0.134829	1.095	0.273482	
## factor(street)235	-0.027998	0.174424	-0.161	0.872496	
## factor(street)236	0.097735	0.123788	0.790	0.429925	
## factor(street)238	-0.001159	0.145275	-0.008	0.993633	
## factor(street)240	-0.205095	0.174265	-1.177	0.239419	
## factor(street)241	-0.019687	0.188805	-0.104	0.916968	
## factor(street)242	0.147439	0.210582	0.700	0.483944	
## factor(street)243	0.243566	0.188984	1.289	0.197662	
## factor(street)244	0.206016	0.130374	1.580	0.114277	
## factor(street)246	-0.033340	0.164319	-0.203	0.839242	
## factor(street)247	0.097857	0.132360	0.739	0.459828	
## factor(street)248	0.100251	0.116940	0.857	0.391429	

```
## factor(street)249      0.189483    0.126718    1.495 0.135047
## factor(street)250      0.148985    0.111842    1.332 0.183033
## factor(street)253      0.220047    0.125137    1.758 0.078878 .
## factor(street)254      0.214999    0.156585    1.373 0.169944
## factor(street)255      0.135722    0.128419    1.057 0.290745
## factor(street)256      0.135936    0.113238    1.200 0.230160
## factor(street)257     -0.003260    0.123801   -0.026 0.978999
## factor(street)258      0.193537    0.120261    1.609 0.107764
## factor(street)259      0.826600    0.210888    3.920 9.28e-05 ***
## factor(street)260      0.067239    0.112486    0.598 0.550095
## factor(street)261      0.061868    0.150351    0.411 0.680771
## factor(street)262      0.148463    0.145378    1.021 0.307315
## factor(street)263      0.128562    0.125097    1.028 0.304261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3229 on 1476 degrees of freedom
## Multiple R-squared:  0.4451, Adjusted R-squared:  0.3759
## F-statistic: 6.434 on 184 and 1476 DF,  p-value: < 2.2e-16

paste("The treatment effect estimate is", signif(summary(model4h)$coefficients[2,1],3))

## [1] "The treatment effect estimate is 0.106"

paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.026"

paste("The 95% robust confidence interval is (", ans[1],",",ans[2],")")

## [1] "The 95% robust confidence interval is ( 0.0547 , 0.157 )"
```

Adam Yang: The 95% robust confidence interval for bins without stickers is (0.0547, 0.157) which is less tight compared to the confidence interval we got from part (g), which is (0.0785, 0.155). The confidence intervals differ because one represents all households who received bins while the other represents only the households that received bins without stickers on them. It is possible that the confidence interval got wider because we introduced more variability along with more separate treatment effects. This increases the standard error which widens the confidence interval.

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
```

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?

```
model5a <- lm(vomiting_day14 ~ treat_zmapp, data = d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model5a, row = 2, z_val = 1.96)
# Use coeftest to find the robust p value
coeftestvalues <- coeftest(model5a, vcovHC(model5a))

summary(model5a)

##
## Call:
## lm(formula = vomiting_day14 ~ treat_zmapp, data = d)
##
## 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

paste("The treatment effect estimate is", signif(summary(model5a)$coefficients[2,1],3))

## [1] "The treatment effect estimate is -0.238"

paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0915"

paste("The 95% robust confidence interval is (", ans[1],",",ans[2],",)")

## [1] "The 95% robust confidence interval is ( -0.417 , -0.0584 )"
```

```

paste("The associated p-value is", signif(summary(model5a)$coefficients[2,4],3))

## [1] "The associated p-value is 0.0066"
paste("The associated robust p-value is", signif(coeftestvalues[2,4],3))

## [1] "The associated robust p-value is 0.0108"

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.

model5b <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0, data = d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model5b, row = 2, z_val = 1.96)
# Use coeftest to find the robust p value
coeftestvalues <- coeftest(model5b, vcovHC(model5b))

summary(model5b)

##
## Call:
## lm(formula = vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
##     data = d)
##
## 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

paste("The treatment effect estimate is", signif(summary(model5b)$coefficients[2,1],3))

## [1] "The treatment effect estimate is -0.166"
paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.082"
paste("The 95% robust confidence interval is (", ans[1],",",ans[2],")")

## [1] "The 95% robust confidence interval is ( -0.326 , -0.00486 )"
paste("The associated p-value is", signif(summary(model5b)$coefficients[2,4],3))

## [1] "The associated p-value is 0.0311"

```

```
paste("The associated robust p-value is", signif(coeftestvalues[2,4],3))
```

```
## [1] "The associated robust p-value is 0.0462"
```

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

Adam Yang: I prefer the estimate of the ATE reported in part b. The robust standard error in part b is 0.082, which is smaller than the robust standard error in part a (0.0915). Hence the robust confidence interval for part b is also narrower. It seems like adding the relevant covariates to the regression was able to absorb some of the residual variance. Furthermore, the magnitude of the estimated treatment effect dropped from -0.238 to -0.166. I believe that `vomiting_day0` and `temperature_day0` gives us a sense of how sick the patient is before the experiment which would intuitively have an impact on `vomiting_day14`. Therefore, the drop in magnitude of the estimated ATE might be due to us addressing that omitted variable bias by adding in those covariates into the regression of part b. Lastly, the adjusted R-squared in part b is 0.2895 which is quite a bit larger than the adjusted R-squared in part a (0.06343). This suggests that the independent variables in part b better explains the variation in the dependent 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.

```
model5d <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 + temperature_day14, data = d)
```

```
# Use previously defined equations to find robust confidence intervals
```

```
ans <- processrobust(model5d, row = 2, z_val = 1.96)
```

```
# Use coeftest to find the robust p value
```

```
coeftestvalues <- coeftest(model5d, vcovHC(model5d))
```

```
summary(model5d)
```

```
##
```

```
## Call:
```

```
## lm(formula = vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 +  
##     temperature_day14, data = d)
```

```
##
```

```
## 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
```

```
paste("The treatment effect estimate is", signif(summary(model5d)$coefficients[2,1],3))
```

```
## [1] "The treatment effect estimate is -0.12"
```

```

paste("The robust standard error is", ans[3])

## [1] "The robust standard error is 0.0858"
paste("The 95% robust confidence interval is (", ans[1], ",", ans[2], ")")

## [1] "The 95% robust confidence interval is ( -0.288 , 0.0481 )"
paste("The associated p-value is", signif(summary(model5d)$coefficients[2,4],3))

## [1] "The associated p-value is 0.125"
paste("The associated robust p-value is", signif(coeftestvalues[2,4],3))

## [1] "The associated robust p-value is 0.165"

```

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

Adam Yang: I prefer the estimate of the ATE reported in part b because the robust standard error in part d is larger than that of part b ($0.0858 > 0.082$). Furthermore, I believe that `temperature_day14` might be a post treatment variable because it was measured after the treatment was given, which makes it a bad control. This is because your temperature on day 14 might very much be related to or a consequence of whether or not you are vomiting on day 14. I don't think it is good practice to add another dependent variable into the regression as a covariate because it will introduce a non-zero bias that we cannot know the size or direction of.

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?

```

model5f <- lm(temperature_day14 ~ treat_zmapp + treat_zmapp*male + male
              + vomiting_day0 + temperature_day0, data = d)

# Use previously defined equations to find robust confidence intervals
ans <- processrobust(model5f, row = 2, z_val = 1.96)
# Use coeftest to find the robust p value
coeftestvalues <- coeftest(model5f, vcovHC(model5f))

summary(model5f)

##
## Call:
## lm(formula = temperature_day14 ~ treat_zmapp + treat_zmapp *
##     male + male + vomiting_day0 + temperature_day0, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70157 -0.37725 -0.02702  0.34687  0.73968
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   48.71269    9.26618   5.257 9.14e-07 ***
## treat_zmapp    -0.23087    0.11871  -1.945  0.0548 .
## male           3.08549    0.12644  24.403 < 2e-16 ***
## vomiting_day0  0.04113    0.18208   0.226  0.8218
## temperature_day0 0.50480    0.09508   5.309 7.34e-07 ***
## treat_zmapp:male -2.07669    0.19164 -10.836 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```
##
## Residual standard error: 0.4518 on 94 degrees of freedom
## Multiple R-squared:  0.9059, Adjusted R-squared:  0.9009
## F-statistic: 181 on 5 and 94 DF,  p-value: < 2.2e-16
paste("The estimated treatment effect on females only is", signif(summary(model5f)$coefficients[2,1],3))

## [1] "The estimated treatment effect on females only is -0.231"
paste("The estimated treatment effect on males only is", signif(summary(model5f)$coefficients[2,1] + su

## [1] "The estimated treatment effect on males only is -2.31"
```

Adam Yang: I added an interaction term of *treat_zmapp*male* which means that the estimated treatment effect of ZMapp on day 14 temperature is now $-0.231 + -2.08 * Male$. When the subject is female, the estimated treatment effect of ZMapp on day 14 temperature is $-0.231 + -2.08 * 0 = -0.231$. When the subject is male, the estimated treatment effect of ZMapp on day 14 temperature is $-0.231 + -2.08 * 1 = -2.31$. This shows that the estimated treatment effect of ZMapp on males is quite a bit higher than the estimated treatment effect of ZMapp on females. Specifically, ZMapp reduces men's temperature 2 degrees more than it does for women. Furthermore, according to the summary chart shown above, the estimated treatment effect of ZMapp on the day 14 temperature of women is not statistically significant at a level of 0.05. On the other hand, the heterogeneous treatment effect represented by the interaction term is shown to be statistically significant (p value at interaction term is approximately 0). That means we can reject the null hypothesis that the treatment effect of ZMapp is equal for both men and women.

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

Adam Yang: I would not have much confidence in this finding because my colleague seemed to be fishing for a statistically significant heterogeneous effect, thereby introducing the multiple comparisons problem. He ended up running 20,000,000 regressions with 10,000 different covariates on 2,000 different indicators of health. By running so many hypothesis tests to look for a significant heterogeneous effect, it is very likely that one out of these 20,000,000 hypothesis tests would result with a statistically significant heterogeneous effect. In fact, if we use the equation shown in page 300 of *Field Experiments*, we see that the probability is essentially 1. However, this can be caused by random chance rather than actually being statistically significant. Because he ran 20,000,000 hypothesis tests, he should have applied the Bonferroni correction to his significance level so that the p-value must be smaller than $2.5e-09$ to be considered statistically significant. The p-value we found in part f is $2.2e-16$ which is smaller than $2.5e-09$ so maybe we would believe this result. However, this is a hypothetical situation where we did not run the regression in part f, so I would maintain that I would not have confidence in this finding from just hearing of my colleague's work. I would recommend using a different data sample to test this hypothesis again.

```
# Bonferroni Correction
alpha = 0.05 # significance level of 0.05
m = 20000000 # number of hypotheses being tested
```

```
paste("The new significance level should be",alpha/m)
```

```
## [1] "The new significance level should be 2.5e-09"
```

```
# The probability of finding at least one covariate that significantly interacts with the treatment at
```

```
paste("The probability of finding at least one covariate that significantly interacts with the treatment
```

```
## [1] "The probability of finding at least one covariate that significantly interacts with the treatment
```

- 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?

Adam Yang: In this case, I would be more inclined to believe that the heterogeneous treatment effect really exists because we did not run a bunch of hypothesis tests to fish for a statistically significant result. Furthermore, in part f, we found that the p-value of the interaction term is $2.2e-16$ which means it very unlikely that the null hypothesis is true.

- 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.*)

Adam Yang: My first thought would be that the research question is not very clear. What does my colleague mean by African descent? Does he mean people who are born in Africa? Does he mean people who have lived in Africa all their lives? Does he mean people who have multiple generations of ancestors who have lived in Africa? Does he mean people with darker skin color? Does he mean African Americans who live in the US but have ancestors from Africa?

Determining the other group to compare these “people of African descent” to is equally, if not more, confusing and convoluted. Furthermore, there is no obvious way to randomly apply treatment. How do we randomly assign people to be of “African descent”? Perhaps if you can get the definition of “African descent” clear as well as specify which other group to compare them to, then you can possibly do the analysis with observational data. However, designing an experiment would be very difficult, if not impossible to do. I think the question proposed by my colleague is a fundamentally unidentified question that cannot be answered by any experiment.