PS3 Solution Set

Alex & Daniel

2018

```
## load packages
library(data.table)
library(multiwayvcov)
library(sandwich)
library(lmtest)
library(stargazer)
library(lme4)
```

```
robust.ci <- function(model, se.type = "HC3") {
  require(sandwich)
  if(se.type %in% c("HC1", "HC2", "HC3")) {
    ses <- sqrt(diag(vcovHC(model, type = se.type)))
  } else if(se.type == "cluster") {
    ses <- sqrt(diag(cluster.vcov(model, ~ cluster)))
  }
  df = summary(model)$df[2]
  t.stat = qt(p = 0.975, df = df, lower.tail = TRUE)
  ci.lower = as.numeric(round(coef(model) - 1.96 * ses, 2))
  ci.upper = as.numeric(round(coef(model) + 1.96 * ses, 2))
  return(list(
    point.estimates = coef(model),
    ci = paste0("[", ci.lower, ",", ci.upper, "]")
    ))
}
```

Problem 1

```
#
# Load data.
#

# data.table
fb_dt  <- fread('broockman_green_anon_pooled_fb_users_only.csv')
fb_dt  <- na.omit(fb_dt)
fb_dt1 <- fb_dt[studyno == 1, ]
fb_dt2 <- fb_dt[studyno == 2, ]

# base
fb_df   <- read.csv('broockman_green_anon_pooled_fb_users_only.csv', stringsAsFactors=F)
fb_df   <- fb_df[complete.cases(fb_df$name_recall, fb_df$treat_ad, fb_df$cluster),]
fb_sub1 <- subset(fb_df, studyno == 1)
fb_sub2 <- subset(fb_df, studyno == 2)
```

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.

```
fb_lm1 <- lm(name_recall ~ treat_ad, fb_dt1)
fb_t1  <- summary(fb_lm1)$coef[2,] #coefficients for treatment indicator

robust.ci(model = fb_lm1)
```

```
## $point.estimates
##  (Intercept)     treat_ad
##  0.182468694 -0.009797887
##
## $ci
## [1] "[0.15,0.21]"  "[-0.05,0.03]"
```

```
stargazer(fb_lm1, header=F)
```

Table 1:

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

But, hey, wait a second. That data is distributed binomally. Aren't we supposed to be using logistic regression on it? Well, actulaly, no. Here's the thing, while an OLS might do poorly on a prediction task (predicting outside of the range of a valid probability statement), it *will* do an admirable job at estimating the causal effect. By admirable, I mean to say, unbiased and efficient for the *marginal* effects.

http://politics.as.nyu.edu/docs/IO/2576/pgm2011.pdf

Just watch!

```
fb_norm  <- glm(name_recall ~ treat_ad, fb_dt1, family = "gaussian")
coef(fb_norm)
```

```
##  (Intercept)     treat_ad
##  0.182468694 -0.009797887
```

```
fb_dt1[ , .(m=mean(name_recall)), keyby=treat_ad][ , diff(m)]
```

```
## [1] -0.009797887
```

```r
fb_logit <- glm(name_recall ~ treat_ad, fb_dt1, family = "binomial")
robust.ci(model = fb_lm1)
```

```
## $point.estimates
##  (Intercept)     treat_ad
##  0.182468694 -0.009797887
##
## $ci
## [1] "[0.15,0.21]"  "[-0.05,0.03]"
```

```r
stargazer(fb_lm1, header=F)
```

Table 2:

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

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

Clusters are defined by individuals with the same gender, age, and location. There might be systematic variation between the clusters, introducing additional noise in our estimates. For example, imagine that 24 year old males in one town happen to be much more politically liberal than their cohorts in a neighboring town. If many individuals in that group are placed in treatment it might overstate the treatment effect – or vice versa if not treated. This uncertainty is what raises standard errors.

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

```r
cluster.vcov(fb_lm1, ~ cluster)
```

```
##              (Intercept)      treat_ad
## (Intercept)  0.000341936 -0.0003385450
## treat_ad    -0.000338545  0.0005642347
```

```r
coeftest(fb_lm1, cluster.vcov(fb_lm1, ~ cluster))
```

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

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

```
fb_lm2 <- lm(name_recall ~ treat_ad, fb_dt2)
coeftest(fb_lm2, cluster.vcov(fb_lm2, ~ cluster))
```

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

```
robust.ci(fb_lm2, se.type = "cluster")
```

```
## $point.estimates
##  (Intercept)     treat_ad
##  0.605788423 -0.002803349
##
## $ci
## [1] "[0.57,0.64]"  "[-0.07,0.07]"
```

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?

```
fb_lm0 = lm(name_recall ~ treat_ad, fb_dt)
coeftest(fb_lm0, cluster.vcov(fb_lm0, ~ cluster))
```

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

```
robust.ci(fb_lm0, se.type = "cluster")
```

```
## $point.estimates
## (Intercept)    treat_ad
##   0.4541960  -0.1550732
##
```

```
## $ci
## [1] "[0.42,0.49]"  "[-0.21,-0.1]"
```

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?

```
fb_dt[ , study2 := studyno == 2]
fb_lm00      <- lm(name_recall ~ treat_ad + study2, fb_dt)
fb_lm00_cl   <- coeftest(fb_lm00, cluster.vcov(fb_lm00, ~ cluster))
fb_lm00_cl
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1806848  0.0169702 10.6472   <2e-16 ***
## treat_ad    -0.0067752  0.0204154 -0.3319     0.74
## study2TRUE   0.4260988  0.0206970 20.5875   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
robust.ci(fb_lm00, se.type = "cluster")
```

```
## $point.estimates
##  (Intercept)     treat_ad   study2TRUE
##  0.180684806 -0.006775249  0.426098820
##
## $ci
## [1] "[0.15,0.21]"  "[-0.05,0.03]" "[0.39,0.47]"
```

```
mem <- lmer(name_recall ~ 1 + treat_ad + (treat_ad | study2), fb_dt)
```

```
stargazer(mem, fb_lm00, header=F)
```

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

The result in part (e) is biased. Each study allocated a different share of subjects to treatment. So, differences between the units in "treatment" and "control" in the pooled regression reflect both the actual effect of the treatments and compositional differences between the units in Study 1 and Study 2 – the "pooled treatment group" over-represents Study 1 subjects. When we include a dummy variable for the study, this compositional difference is taken away and regression sees that, within each study, the average treatment effect is indistinguishable from zero.

If we wanted a clear sense of whether the studies are, in fact, different, we might look at an interaction of the study indicator with the rest of the model; in the simplest case, this will just interact the study indicator with the treatment indicator which will give us a clear test for a difference in the treatment effect in different studies.

```
m <- lm(name_recall ~ treat_ad + study2 + treat_ad*study2, fb_dt)
stargazer(m, header=F)
```

Table 3:

	Dependent variable:	
	name_recall	
	linear mixed-effects	*OLS*
	(1)	(2)
treat_ad	−0.007	−0.007
	(0.019)	(0.018)
study2		0.426***
		(0.018)
Constant	0.394*	0.181***
	(0.212)	(0.016)
Observations	2,701	2,701
R^2		0.193
Adjusted R^2		0.193
Log Likelihood	−1,612.673	
Akaike Inf. Crit.	3,237.345	
Bayesian Inf. Crit.	3,272.754	
Residual Std. Error		0.438 (df = 2698)
F Statistic		322.848*** (df = 2; 2698)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4:

	Dependent variable:
	name_recall
treat_ad	−0.010
	(0.024)
study2	0.423***
	(0.023)
treat_ad:study2	0.007
	(0.037)
Constant	0.182***
	(0.019)
Observations	2,701
R^2	0.193
Adjusted R^2	0.192
Residual Std. Error	0.438 (df = 2697)
F Statistic	215.167*** (df = 3; 2697)
Note:	*p<0.1; **p<0.05; ***p<0.01

So, there is no evidence for a different treatment effect, but a clear difference due to compositional differences in the studies!

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."

Political campaigns may choose to run ads in the kinds of places where voters tend to be most supportive of their causes already (for example, if the ads remind people to vote in the election, campaigns might be more interesting in reminding those who agree with them than those who will vote against them). As a result, differences between vote returns in the areas where campaigns run ads and do not run them may reflect existing differences, not the effect of the advertising.

"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."

The kinds of people who see online ads may be quite different than the kinds of people who do not. For example, young people and wealthier people may have different kinds of opinions than their counterparts. Likewise, those who use the internet more may be exposed to other kinds of media content as well - for example, they have read newspapers that contain views friendly to the campaign.

It also seems like Facebook targeted these ads to those who were likely to be most supportive already – so, if we find that those exposed to the ads were more supportive, this might just reflect pre-existing differences.

Problem 2

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.

$$\begin{aligned} ATE &= 0.187 \\ CI &= 0.187 \pm 1.96 * 0.032 \\ &= (0.1243, 0.2497) \end{aligned}$$

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.

$$\begin{aligned} ATE &= -0.024 \\ CI &= -0.024 \pm 1.96 * 0.039 \\ &= (-0.1004, 0.0524) \end{aligned}$$

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

1, 2, 3, and 4 (% of visits turned in bag, avg no of bins turned in per week, avg weight of recyclables turned in per week, avg market value of recyclables given per week). All but avg % of contamination per week, outcome number 5.

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

None.

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

Average difference $= 0.281 * 2 = 0.562$. We multiply by 2 because the problem states that Household A turned in 2 kg more.

f. Suppose that the variable "percentage of visits turned in bag, baseline" had been left out of the regression reported in Column 1. What would you expect to happen to the results on providing a recycling bin? Would you expect an increase or decrease in the estimated ATE? Would you expect an increase or decrease in the standard error? Explain your reasoning.

Assuming there is no selection bias that favors putting households with particularly high baseline recycling percentages in the treatment or control groups (as we should be able to assume because of the randomization), then we would expect no change in the ATE. The reason we don't expect the ATE to change in the absence of selection bias, is that the baseline level should be independent of treatment with proper randomization. (The ATE may move around a bit due to covariate adjustment, so it probably won't literally stay the same, but it shouldn't change much.)

However, we would expect the standard error to increase. By removing that baseline from the model, it's harder to tell statistically what variation in outcomes is being driven by treatment versus just being a household that recycles a lot, thus the increase in standard errors.

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

No, it's not a bad control because the treatments probably don't cause people to get a cell phone. Regardless of whether the researchers did this experiment or not, the subjects with cell phones would still have cell phones and the subjects without cell phones would not.

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.

It probably would not change very much because having a cell phone isn't likely to be correlated with treatment. If having a cell phone is highly prognostic of the outcome in general, however, it may improve the estimate.

Most importantly, *subsetting* the data and only running the regression among those who do have a cell phone would probably produce a more precise estimate of the effect, since there probably isn't any effect of an SMS message among those without cell phones.

Problem 3

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.)

3x3 (3 types of bins: bin with sticker, bin without sticker, no bin; and 3 types of SMS messages: personal, generic, none). However, I would understand if you did something strange with the people who don't have a cell phone. This is a little wrinkle to the easy categorization of this design.

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?

No bins, no cell phone, and no SMS message.

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

Regardless of the subjects' other characteristics, giving them a bin without a sticker on average increased the percentage visits on which that subject turned in a bag by 3.5 percentage points.

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

The bin with the sticker appears to have a stronger treatment effect, causing a 5.5 percentage point increase in "percentage of visits turned in bag" compared to 3.5 percentage points for bins without stickers, almost 60% larger.

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

That difference is not statistically significant as indicated by the F-Test p-value of 0.31.

f. Notice that Table 4C is described as results from "fully saturated" models. What does this mean? Looking at the list of variables in the table, explain in what sense the model is "saturated."

Fully saturated models use dummy variables to capture all treatments and covariates. In this case that means creating a dummy variable for every possible combination of Bins and SMS messages.

Problem 4

```
# Read in data.
library(foreign)
karlan <- read.dta('karlan_data_subset_for_class.dta', convert.underscore=TRUE)
karlan <- data.table(karlan)
```

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.

```
lma <- lm(avg.bins.treat ~ bin, data = karlan)
stargazer(lma, header=F)
```

Table 5:

	Dependent variable:
	avg.bins.treat
bin	0.135***
	(0.020)
Constant	0.635***
	(0.012)
Observations	1,785
R^2	0.024
Adjusted R^2	0.024
Residual Std. Error	0.405 (df = 1783)
F Statistic	44.516*** (df = 1; 1783)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
coeftest(lma, vcovHC(lma))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.635350   0.011477 55.3599 < 2.2e-16 ***
## bin         0.135380   0.020823  6.5015  1.03e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(lma)
```

```
##                   2.5 %    97.5 %
## (Intercept) 0.61221964 0.6584797
## bin         0.09558421 0.1751758
```

```
robust.ci(lma, se.type = "HC3")
```

```
## $point.estimates
## (Intercept)         bin
##   0.6353497   0.1353800
##
## $ci
## [1] "[0.61,0.66]" "[0.09,0.18]"
```

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.

```
lmb <- lm(avg.bins.treat ~ bin + base.avg.bins.treat, data = karlan)
stargazer(lmb,header=F)
```

Table 6:

	Dependent variable:
	avg.bins.treat
bin	0.125^{***}
	(0.017)
base.avg.bins.treat	0.393^{***}
	(0.013)
Constant	0.350^{***}
	(0.014)
Observations	1,785
R^2	0.342
Adjusted R^2	0.342
Residual Std. Error	0.333 (df = 1782)
F Statistic	463.891^{***} (df = 2; 1782)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
coeftest(lmb, vcovHC(lmb))
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value  Pr(>|t|)
## (Intercept)          0.349603   0.021163 16.5192 < 2.2e-16 ***
## bin                  0.124693   0.017193  7.2527 6.066e-13 ***
## base.avg.bins.treat 0.392965   0.030337 12.9533 < 2.2e-16 ***
```

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

```
confint(lmb)
```

```
##                           2.5 %    97.5 %
## (Intercept)          0.32267158 0.3765335
## bin                  0.09200378 0.1573822
## base.avg.bins.treat 0.36671039 0.4192189
```

```
robust.ci(lmb, se.type = "HC3")
```

```
## $point.estimates
##         (Intercept)                 bin base.avg.bins.treat
##           0.3496026           0.1246930           0.3929647
##
## $ci
## [1] "[0.31,0.39]" "[0.09,0.16]" "[0.33,0.45]"
```

The confidence interval here is slightly narrower than in the previous subsection because the pre-treatment value of Y is predictive of Y, allowing us to reduce noise when we include it in the regression.

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.

```
karlan[ , f.street := as.factor(street)]
```

```
k.mod1 <- lm(avg.bins.treat ~ bin + base.avg.bins.treat + f.street, data = karlan)
stargazer(k.mod1, type = "latex", omit = "f.street", header=F) # we don't actulaly care about these mea
```

Table 7:

	Dependent variable:
	avg.bins.treat
bin	0.114***
	(0.017)
base.avg.bins.treat	0.374***
	(0.014)
Constant	0.368***
	(0.032)
Observations	1,782
R^2	0.436
Adjusted R^2	0.372
Residual Std. Error	0.324 (df = 1600)
F Statistic	6.840*** (df = 181; 1600)
Note:	*p<0.1; **p<0.05; ***p<0.01

Let's pause here to think about what the fixed effects are telling us. Do we *actually* care about the average outcome at the street level? No! Or, at least, I don't. What I *do* care about is that I am "soaking up" some street level variance in the Y variable. Basically, by including the street level fixed effect, I am removing

differences that I *could* have otherwise modeled, but chose not to. For example, I might (reasonably) think that there are differences in the home values on some streets relative to others. Then, I *could* have gone to Zillow (for Peru), scraped and munged the data about the home values on each of the streets, and then included that `average.street.home.price` variable that I made in the regression. **OR**, and this is easier, I can just say, "Meh, I think there are differences between the streets, let's just average that out."

```
stargazer(lmb, k.mod1, type = "latex", omit = "street", header = FALSE,
          add.lines = list(c("Robust SEs", "No", "Yes")))
```

Table 8:

	Dependent variable:	
	avg.bins.treat	
	(1)	(2)
bin	0.125***	0.114***
	(0.017)	(0.017)
base.avg.bins.treat	0.393***	0.374***
	(0.013)	(0.014)
Constant	0.350***	0.368***
	(0.014)	(0.032)
Robust SEs	No	Yes
Observations	1,785	1,782
R^2	0.342	0.436
Adjusted R^2	0.342	0.372
Residual Std. Error	0.333 (df = 1782)	0.324 (df = 1600)
F Statistic	463.891*** (df = 2; 1782)	6.840*** (df = 181; 1600)
Note:		*p<0.1; **p<0.05; ***p<0.01

```
# prevent all the fixed effects for
# each street from printing out
robust.ci(k.mod1, se.type = "HC3")$point.estimates[1:3]
```

```
##         (Intercept)                 bin base.avg.bins.treat
##           0.3677440           0.1138868           0.3737068
```

```
robust.ci(k.mod1, se.type = "HC3")$ci[1:3]
```

```
## [1] "[0.3,0.44]"  "[0.08,0.15]" "[0.31,0.43]"
```

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.

Conducting blocked random assignment increases the precision with which the treatment effect is estimated to the extent that the blocks help predict the outcome. In this case, unlike in part b, it seems like the blocks probably didn't predict the outcome very well because they did not decrease our uncertainty very much.

Suppose the blocks did predict the outcome really well. If we didn't include block indicators in the regression, our estimated standard error would fail to get "credit" for the reducing the noise in the ATE. The fact that

the block indicators fail to reduce the standard error means that the blocking turned out not to have much value.

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

```
karlan[ , no.cell := 1 - havecell]
```

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

```
lmf <- lm(avg.bins.treat ~ bin + base.avg.bins.treat + no.cell + f.street,
          data = karlan)
stargazer(lmf, type = "latex", omit = "street", header = FALSE)
```

Table 9:

	Dependent variable:
	avg.bins.treat
bin	0.115^{***}
	(0.017)
base.avg.bins.treat	0.373^{***}
	(0.014)
no.cell	-0.050^{***}
	(0.017)
Constant	0.387^{***}
	(0.032)
Observations	1,781
R^2	0.439
Adjusted R^2	0.375
Residual Std. Error	0.323 (df = 1598)
F Statistic	6.875^{***} (df = 182; 1598)
Note:	*p<0.1; **p<0.05; ***p<0.01

Whether or not subjects have a cell phone does not seem to predict Y well, meaning that it does not change the standard error on the treatment effect coefficient *very* much. Yes, it does change it somewhat, and so if you interpreted the change as being a substantial one, we won't hold it against you. This is sort of a judgment call on your part whether the change is a lot or a little. Relative to baseline, I would call this a small change.

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.

```
lmg <- lm(avg.bins.treat ~ bin + base.avg.bins.treat + no.cell +
                    sms + f.street, data = karlan)
stargazer(lmg, type = "latex", omit = "street", header = FALSE)
```

Table 10:

	Dependent variable:
	avg.bins.treat
bin	0.115^{***}
	(0.017)
base.avg.bins.treat	0.373^{***}
	(0.014)
no.cell	-0.047^{**}
	(0.020)
sms	0.005
	(0.021)
Constant	0.385^{***}
	(0.034)
Observations	1,781
R^2	0.439
Adjusted R^2	0.375
Residual Std. Error	0.323 (df = 1597)
F Statistic	6.834^{***} (df = 183; 1597)
Note:	*p<0.1; **p<0.05; ***p<0.01

The SMS treatment is random, so we know it can't predict Y very well. It might have a bit of an effect on Y, but in general we're going to expect similar values of Y across randomly assigned treatment groups. Therefore, it doesn't increase our precision very 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.

```
lmh <- lm(avg.bins.treat ~ sms.p + sms.g + bin.s + bin.g +
            base.avg.bins.treat + no.cell + f.street, data = karlan)
stargazer(lmh, type = "latex", omit = "street", header = FALSE)
```

The confidence interval is wider. This is because the number of subjects who were assigned to just the unadorned bin is lower than the number of subjects who were assigned to any bin, so the sample size for this

Table 11:

	Dependent variable:
	avg.bins.treat
sms.p	-0.008
	(0.025)
sms.g	0.020
	(0.025)
bin.s	0.128^{***}
	(0.022)
bin.g	0.103^{***}
	(0.022)
base.avg.bins.treat	0.374^{***}
	(0.014)
no.cell	-0.046^{**}
	(0.020)
Constant	0.385^{***}
	(0.034)
Observations	1,781
R^2	0.440
Adjusted R^2	0.375
Residual Std. Error	0.323 (df = 1595)
F Statistic	6.769^{***} (df = 185; 1595)
Note:	*p<0.1; **p<0.05; ***p<0.01

comparison is lower.

Problem 5 - Ebola

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?

```
ebola <- read.csv('ebola_rct2.csv')
lma      <- lm(vomiting_day14 ~ treat_zmapp, ebola)
coeftest(lma, vcov = vcovHC(lma, "HC3"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.847458   0.047616  17.798  < 2e-16 ***
## treat_zmapp -0.237702   0.091459  -2.599  0.01079 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is 0.0066.

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.

```
lmb <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0, ebola)
robust.ci(lmb, se.type = "HC3")
```

```
## $point.estimates
##      (Intercept)      treat_zmapp    vomiting_day0 temperature_day0
##     -19.46965517      -0.16553674       0.06455724       0.20554815
##
## $ci
## [1] "[-34.38,-4.56]" "[-0.33,0]"      "[-0.28,0.41]"   "[0.05,0.36]"
```

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

The estimate in part (b) happens to be a bit smaller, but it's also more precise. Therefore, the estimate in part (b) is likelier to be closer to accurate. In general, we would prefer the estimate in part (b).

On the other hand, there are also principled reasons to prefer part (a). For example, one might worry that the covariates selected reflected a fishing expedition.

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.

```
lmd <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0 +
           temperature_day14, ebola)
robust.ci(lmd, se.type = "HC3")
```

```
## $point.estimates
##       (Intercept)       treat_zmapp     vomiting_day0  temperature_day0
##      -22.59158542       -0.12010063        0.04603820        0.17664160
## temperature_day14
##        0.06014826
##
## $ci
## [1] "[-37.77,-7.41]" "[-0.29,0.05]"   "[-0.29,0.39]"   "[0.03,0.33]"
## [5] "[0.01,0.11]"
```

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

The estimate from part (b). The estimate in part (d) is biased because temperature on day 14 is a "bad control" – if the treatment works, it may have an impact on temperature on day 14.

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?

```
lmf <- lm(temperature_day14 ~ treat_zmapp * male + vomiting_day0 +
           temperature_day0, ebola)
robust.ci(lmf, se.type = "HC3")
```

```
## $point.estimates
##      (Intercept)      treat_zmapp             male    vomiting_day0
##      48.71268983      -0.23086555       3.08548611        0.04113066
## temperature_day0 treat_zmapp:male
##       0.50479728      -2.07668626
##
## $ci
## [1] "[28.73,68.69]" "[-0.46,0]"     "[2.85,3.32]"   "[-0.34,0.42]"
## [5] "[0.3,0.71]"    "[-2.47,-1.69]"
```

The results suggest the treatment works especially well for males. This is indicated on the *hightly* significant effect on the interaction term for males.

HERE IS SOME ADDITIONAL CANDY

We have talked in class that one way to present this work to stakeholders would be to *test* using an interaction model, but to report it to folks who don't want to think through the interaction as a subset model. Here, I'm presenting two ways of reporting this for you. First, the subsetting.

```
lmf.all  <- lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
               data = ebola)
lmf.male <- lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
               data = ebola[ebola$male == 1, ]) # note that we're using data.frames here, not data.table
lmf.female <- lm(temperature_day14 ~ treat_zmapp + vomiting_day0 + temperature_day0,
               data = ebola[ebola$male == 0, ])
```

```
stargazer(lmf.all, lmf.male, lmf.female, lmf,
          type = "latex",
          add.lines = list(c("Subset", "All", "Males", "Femles", "All")),
          header = FALSE,
          title = "Comparison of Estimates in Subset and Interaction Models",
          covariate.labels = c("Treatment", "Male", "Baseline Vomiting", "Baseline Temp", "Male * Treat
          column.sep.width = "0pt",
          omit.stat = "all"
          )
```

Table 12: Comparison of Estimates in Subset and Interaction Models

	Dependent variable:			
	temperature_day14			
	(1)	(2)	(3)	(4)
Treatment	-0.755^{***}	-2.239^{***}	-0.229^{*}	-0.231^{*}
	(0.259)	(0.154)	(0.119)	(0.119)
Male				3.085^{***}
				(0.126)
Baseline Vomiting	0.308	−0.379	0.265	0.041
	(0.500)	(0.297)	(0.227)	(0.182)
Baseline Temp	0.481^{*}	0.767^{***}	0.374^{***}	0.505^{***}
	(0.261)	(0.161)	(0.116)	(0.095)
Male * Treat Interaction				-2.077^{***}
				(0.192)
Constant	51.904^{**}	26.232	61.463^{***}	48.713^{***}
	(25.435)	(15.698)	(11.341)	(9.266)
Subset	All	Males	Femles	All
Note:				*p<0.1; **p<0.05; ***p<0.01

A second way to deal with this is to *predict* the outcome back out into a space that is readily interpertable. This might take the following form:

1. Fit a model with an interaction effect
2. Create a new dataset that creates the contrasts that you would like to highlight.
3. Use the `predict` function to predict values in the Y that "make sense".

```
n.data <- data.frame(treat_zmapp = c(1,0,1,0),
                     male = c(1,1,0,0),
                     vomiting_day0 = mean(ebola$vomiting_day0),
                     temperature_day0 = mean(ebola$temperature_day0)
                    )

p <- predict(lmf, type = "response", newdata = n.data, se.fit = TRUE)
```

Here, the predicted change for males is 2.3075518, while the predicted change for the females is 0.2308655.

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

Because my colleague went on a huge fishing expedition, I would be pretty skeptical that ZMapp actually works especially well for reducing temperature in men. By checking so many different heterogeneous treatment effects, it was extremely likely that my colleague would find a big one by chance, even if there were none.

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?

Yes. If this were the first hypothesis I checked, I'd be quite confident in rejecting the null hypothesis that the effects are the same for men and women, as the p-value is almost zero! When just checking one hypothesis, p-values retain their interpretation as "the likelihood I'd get this data if I were wrong." What's important is, the very same data should lead us to conclude different things if we looked for this hypothesis right away or if we only found it after a big fishing expedition.

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.)

Our colleague is proposing a Fundamentally Unidentified Question. The question is both ambiguous and not causal. It may be true that people of African descent are likelier to get Ebola. But to say "being of African descent causes ebola risk" is to make a claim about an implicit manipulation of some kind, which just doesn't make sense here – we can't randomly assign people's heritage.