Field Experiments - PS4

Ray Buhr

July 20, 2015

library(dplyr)

##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:stats':  
##   
## filter, lag  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

#### 1. FE exercise 5.2.

**a. In addition, please also answer this question: Which population is more relevant to study for future decisionmaking: the set of Compliers, or the set of Compliers plus NeverTakers? Why?**

In a scenario where the ATE was positive, but the CACE was negative, the CACE would be useful in comparing only the same kinds of people and not useful when trying to measure the net effect of the treatment. For example, testing whether justice meetings (meetings between the families of victims and the criminal) a negative CACE would be useful to understand how that meeting would actually benefit the families who were able to take part in the meeting, but would not accurately represent the overall costs of scheduling the meetings or the toll of non-compliance if the criminal could/did not attend. Deciding which population to study for future decision making depends on which factors are most important to the core question of the study. In the justice meeting example, the core question is whether creating an environment where those meetings are schedule has greater net benefit than net costs, so the ATE would be more informative.

#### 2. FE exercise 5.6.

1000 voters in treatment, 2000 in control  
canvasers state they contacted 500 voters in treatment, but actually only reached 250  
400/1000 in treatment actually voted  
700/2000 in control actually voted

**a.) estimate CACE with belief that 500 treatment received contact**

CACE = ATE / compliers  
= (400/1000 - 700/2000) / (500/1000)  
= 0.1

**b.) estimate CACE with knowledge that only 250 in treatment received contact**

= (400/1000 - 700/2000) / (250/1000)  
= 0.2

**c.)**

The canvasers exaggerated contact numbers underestimate the effectiveness of actually making contact. Looking at the CACE, the effect of making contact based on the exaggerated contact rate equates to a 50% smaller effect than actually happened.

#### 3. FE exercise 5.10.

library(foreign)  
guan\_green <- read.dta("~/MIDS/241/Guan\_Green\_CPS\_2006.dta")

**a.) estimate the ITT**

library(lmtest)

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

library(sandwich)  
guan\_green <- na.omit(guan\_green)  
itt\_fit <- lm(turnout ~ treat2, data = guan\_green)  
coeftest(itt\_fit, vcovHC(itt\_fit))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.668666 0.012897 51.8470 < 2.2e-16 \*\*\*  
## treat2 0.131930 0.015025 8.7804 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The ITT is 0.1319296

**b.) Test the sharp null hypothesis that ITT = 0 considering random assignment was clustered by dorm room.**

gg\_lm <- lm(turnout ~ treat2 + dormid, data=guan\_green)  
cl(gg\_lm, guan\_green$dormid)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.3382e-01 2.7046e-02 23.4347 < 2.2e-16 \*\*\*  
## treat2 1.3273e-01 2.3239e-02 5.7113 1.202e-08 \*\*\*  
## dormid 3.3576e-09 1.6830e-09 1.9949 0.04612 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

At the 95% confidence level, with a standard error of 0.0232393 and a p-value of 1.201940810^{-8}, we find evidence that does not support the sharp null hypothesis that ITT = 0.

**c.) estimate CACE**

library(AER)

## Loading required package: car  
## Loading required package: survival

cace\_fit <- ivreg(turnout ~ treat2, data=guan\_green, ~contact)   
coeftest(cace\_fit, vcovHC(cace\_fit))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.661233 0.014157 46.707 < 2.2e-16 \*\*\*  
## treat2 0.143051 0.017235 8.300 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

CACE is 0.1430509

**d.) Write down a model of the expected turnout rates in the treatment and control groups, incorporating the average effect of the leaflet.**

expected turn out rate (ETR) = E[Yi(z=1, d=1)] \* treated + 1% \* E[Yi(z=1, d=0)] (leaflets) + E[Y\_i(z=0, d=0)] \* control  
effect of treated E[Yi(z=1, d=1)] = control E[Y\_i(z=0, d=0)] + leaflet + CACE  
effect on non-compliers in treatment E[Yi(z=1, d=0)] = control E[Yi(z=0, d=0)] + leaflet

**e.) given this assumption, estime CACE**

effect of treated = 892/1334 + 0.01 + CACE = 0.679 + CACE  
effect on non-compliers in treatment = 892/1334 + 0.01 = 0.679  
= (0.679 + CACE) \* treated + 0.679 \* leaflets + 0.669 \* control = 1616.02 + 2380 \* CACE + 209.132 + 892  
= 2380 \* CACE + 2717.152  
(2152 + 892) = 2717.152 + 2380 \* CACE  
CACE = 0.1373311

**f.) assume leaflet had no effect on compliers, but increased never-takers by 3%, estimate CACE**

effect of treated = 892/1334 + 0.00 + CACE = 0.669 + CACE  
effect on non-compliers in treatment = 892/1334 + 0.03 = 0.699  
= (0.669 + CACE) \* treated + 0.699 \* leaflets + 0.669 \* control = 1592.22 + 2380 \* CACE + 215.292 + 892  
= 2380 \* CACE + 2699.512 (2152 + 892) = 2699.512 + 2380 \* CACE  
CACE = 0.1447429

#### 4. FE exercise 5.11

**a. We are rewriting part (a) as follows: Estimate the proportion of Compliers by using the data on the Treatment group. Then compute a second estimate of the proportion of Compliers by using the data on the Placebo group. Are these sample proportions statistically significantly different from each other? Explain why you would not expect them to be different, given the experimental design.**

nickerson <- data.frame(  
 treated = c(rep(0,2572), rep(1, 486), rep(0, 2086), rep(1, 470), rep(0, 2109)),  
 assigned = c(rep(0,2572), rep(1, 2572), rep(1, 2579)),  
 placebo = c(rep(0,2572), rep(0,2572), rep(1, 2579)),  
 voted = c(rep(1, 803), rep(0, 1769),   
 rep(1, 190), rep(0, 296), rep(1, 683), rep(0, 1403),   
 rep(1, 140), rep(0, 330), rep(1, 678), rep(0, 1431))  
)

Proportion of compliers in treatment group = 486 / (486 + 2086) = 18.896%  
Proportion of compliers in placebo group = 470 / (470 + 2109) = 18.224%  
Different of means = 0.672%

The standard error in values among those assigned treatment and those who complied:

treatment <- nickerson %>% filter(placebo == 0, assigned == 1)  
# standard error = sd / sqrt(n)  
sd(treatment$treated)/sqrt(length(treatment$treated))

## [1] 0.00772064

Since the difference in means (0.0067168) is less than twice the standard error (0.0154413) we can't say with any certainty that the difference is not due to chance. We wouldn't expect the difference to be much different because no matter what the message we would still expect those who could be present to receive contact to accept the message.

**b.) Does the data suggest never-takers in treatment and placebo groups have the same turnout rates?**

The results table shows that the percent of never-takes that turnout to vote for treatment was 32.74% and for placebo was 32.15%. The difference in means here is 0.0059, which at twice the standard error calculated above shows there is not a significant statistical difference between the turnout rates for never-takers in treatment vs. placebo. The comparison shows us the experiment design was solid since the failure of delivery of any message did not have an effect on outcome.

**c.) Estimate the CACE of receiving placebo. Is the estimate consistent with the assumption that placebo has no effect of turnout?**

baseline <- nickerson %>% filter(assigned == 0)  
placebo <- nickerson %>% filter(placebo == 1)  
placebo\_effect <- lm(placebo$voted ~ placebo$assigned)  
summary(placebo\_effect)

##   
## Call:  
## lm(formula = placebo$voted ~ placebo$assigned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.3172 -0.3172 -0.3172 0.6828 0.6828   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.317177 0.009166 34.6 <2e-16 \*\*\*  
## placebo$assigned NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4655 on 2578 degrees of freedom

ITT <- placebo\_effect$coefficients[[1]] - mean(baseline$voted)  
ITT\_d <- mean(placebo$treated)  
CACE <- ITT / ITT\_d

With the CACE = 0.027265 and the standard error as seen in the model summary to be 0.009166, we can calculate the standard error of CACE as 0.009166 / ITT\_d = 0.050296. Since the CACE is even smaller than the CACE SE, we can say there is very likely no effect of the placebo on the voter turnout.

**d.) Estimate the CACE of treatment using 2 different methods. First use conventional method of dividing ITT by ITT\_d. Second compare turnout rates among complieres in treatment and placebo group.**  
*Hint: ITT D means the average effect of the treatment on the dosage of the treatment. I.E., its the contact rate.*

First method:

treatment\_effect <- lm(voted ~ treated, data = treatment)  
ITT <- treatment\_effect$coefficients[[1]] - mean(baseline$voted)  
ITT\_d <- mean(treatment$treated)  
CACE\_1 <- ITT / ITT\_d  
CACE\_1

## [1] 0.08050732

Second method:

treatment\_contact\_rate <- 486/2572  
placebo\_contact\_rate <- 470/2579  
CACE\_2 <- .3909 - .2979  
CACE\_2

## [1] 0.093

The results of comparing the two methods of calculating CACE came back slightly different. The second method has higher statistical precision because we can compare compliers directly thus having higher statistical power (less chance of incorrectly rejecting the null hypothesis).

#### 5. Determine the direction of bias in estimating the ATE for each of the following situations when we randomize at the individual level. Do we overestimate, or underestimate? Briefly but clearly explain your reasoning.

**a.) In the advertising example of Lewis and Reiley (2014), assume some treatment group members are friends with control group members.**

Treatment members who are friends with control members will lead to an understimate of the ATE because the friends who received treatment (saw ad) may talk to their friends about it thus causing some spillover to the control which diminishes the effect that would be observed without the influence from the friends in treatment group.

**b.) Consider the police displacement example from the bulleted list in the introduction to FE 8, where we are estimating the effects of enforcement on crime.**

The displacement of crime from one neighborhood to other nearby neighborhoods would overestimate the effect of the police campaign in the targeted neighborhood. If you could prevent the criminals from spilling over into other nearby areas, then they may still committ some of the crimes they took to other nearby areas since it was easier to get away with there. By not counting the crimes that just moved away a small distance, the effect of the police campaign look be larger than it really was.

**c.) Suppose employees work harder when you experimentally give them compensation that is more generous than they expected, that people feel resentful (and therefore work less hard) when they learn that their compensation is less than others, and that some treatment group members talk to control group members.**

When treatment group members for this study, people receiving boosted compensation, talk to control members, people receiving regular compensation, the effect on the control group would be amplified because they would feel resentful and thus work less hard than if they didn't know the treatment group was earning more, so the bias would be an overestimation.

**d.) When Olken (2007) randomly audits local Indonesian governments for evidence of corruption, suppose control group governments learn that treatment group governments are being randomly audited and assume they are likely to get audited too.**

The reaction of the control group members to the random audit of treatment groups likely causes them to behave in a less corrupt fashion than if they had no idea that other government groups were being audited. Therefore the bias is going to be an underestimate of actual corruption since the control is shifting behavior to hide their corrupt actions.

#### 6. FE exercise 8.2.

Random assignment to roommate pairs ensures that the two people are not self-selecting to room with someone similar to themselves. In the roommate pairs found to have correlated weights at the end of their freshmen year of college, perhaps that is because both people chose to room with someone eats and drinks in a similar style and thus would experience similar changes in weight over the year. In cases where the roommates did not choose each other, perhaps the roommates did not spend as much time together and never mimicked the others consumption habits.

#### 7. FE exercise 8.6.

clinic <- read.csv("~/MIDS/241/GerberGreenBook\_Chapter8\_Table\_8\_2.csv", stringsAsFactors = F)

**a.) Estimate E[Y01 - Y00] for random assignment that places the treatment at location A.**

(0/0.2) / (1/0.2) - ((0/0.4)+(6/0.6)+(6/0.8)) / ((1/0.4)+(1/0.6)+(1/0.8))  
= -3.2307692

**b.) Estimate E[Y10 - Y00] for random assignment that places the treatment at location A, restricting the sample to the set of villages that have a non-zero probability of expressing both of these potential outcomes.**

(2/0.4) / (1/0.4) - ((0/0.4)+(6/0.6)) / ((1/0.4)+(1/0.6))  
= -0.4

**c.) In order to make a more direct comparison between these two treatment effects, estimate E[Y01 - Y00], restricting the sample to the same set of villages in part b.**

(0/0.2) / (1/0.2) - ((0/0.4)+(6/0.6)) / ((1/0.4)+(1/0.6)) = -2.4

#### 8. FE exercise 8.9.

hotspots <- read.csv("~/MIDS/241/GerberGreenBook\_Chapter8\_Table\_8\_4\_8\_5.csv", stringsAsFactors = F)

**a.) For the subset of 11 hotspot locations that lie outside the range of possible spillovers, calculate E[Y01 - Y00], the ATE of immediate police surveillance.**

hotspots\_out <- hotspots[hotspots$hotwitin500==0, ]  
ATE <- mean(hotspots\_out[hotspots\_out$assignment==1,]$y) - mean(hotspots\_out[hotspots\_out$assignment==0,]$y)  
ATE

## [1] 3.333333

**b.) For the remaining 19 hotspot locations that lie within the range of possible spillovers, calculate E[Y01 - Y00], E[Y10 - Y00], and E[Y11 - Y00].**

hotspots\_in <- hotspots[hotspots$hotwitin500!=0, ]

E[Y01 - Y00] =

mean(hotspots\_in[hotspots\_in$assignment==1,]$y) - mean(hotspots\_in[hotspots\_in$assignment==0,]$y)

## [1] -15.01429

E[Y10 - Y00] =

mean(hotspots\_in[hotspots\_in$exposure==10,]$y) - mean(hotspots\_in[hotspots\_in$exposure==0,]$y)

## [1] 5.888889

E[Y11 - Y00] =

mean(hotspots\_in[hotspots\_in$exposure==11,]$y) - mean(hotspots\_in[hotspots\_in$exposure==0,]$y)

## [1] -8.111111

**c.) Use te data to estimate the average effect of spillover on nonexperimentsal units.**

non\_exp <- read.csv("~/MIDS/241/GerberGreenBook\_Chapter8\_Exercise\_9c.csv", stringsAsFactors = F)  
y10 <- non\_exp[non\_exp$exposure==10,]$y  
y00 <- non\_exp[non\_exp$exposure==0,]$y  
y10\_probs <- non\_exp[non\_exp$exposure==10,]$prob10  
y00\_probs <- non\_exp[non\_exp$exposure==0,]$prob00  
sum(y10 / y10\_probs) / sum(1 / y10\_probs)

## [1] 8.88801

sum(y00 / y00\_probs) / sum(1 / y00\_probs)

## [1] 3.710755

The effect of spillover on nonexperimental units is thus:

sum(y10 / y10\_probs) / sum(1 / y10\_probs) - sum(y00 / y00\_probs) / sum(1 / y00\_probs)

## [1] 5.177255

#### 9. FE exercise 8.10.

**a.) Explain the assumptions needed to identify this causal effect based on this within-subjects design. Are these assumptions plausible? What special concerns arise due to the fact that the subject was conducting the study, undergoing the treatments and measuring her own outcomes?**

The assumptions needed to identify causal effects are that walking or running in the morning is truly random, that other aspects of her life (such as eating, drinking, other exercise) stay the same each day, that the difficulty in the tetris game stays the same each day, and that playing tetris every day doesn't make her better or worse at it. Some of the assumptions could be plausible, such as maintain daily caloric intake and restricting other exercise, but the assumption that playing tetris each day doesn't change skill in the game is not plausible. It could be a concern that the subject does not fully comply with all treatments each day, but fails to record such in order to prevent the study from becoming invalid.

**b.) Estimate the effect of running on Tetris score.**

tetris <- read.dta("~/MIDS/241/Hough\_WorkingPaper\_2010.dta")  
run\_effect\_tetris <- lm(tetris$tetris ~ tetris$run)  
summary(run\_effect\_tetris)

##   
## Call:  
## lm(formula = tetris$tetris ~ tetris$run)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19294 -6707 -1154 4890 29628   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12806 3708 3.453 0.00226 \*\*  
## tetris$run 13613 4856 2.804 0.01035 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11730 on 22 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.2632, Adjusted R-squared: 0.2297   
## F-statistic: 7.86 on 1 and 22 DF, p-value: 0.01035

The effect of running on Tetris would be 1.3613110^{4} with a p-value of 0.01035.

**c.) Use variable Run to predict Tetris score on the preceding day. Presumably the true effect is zero. Does randomization inference confirm?**

tetris$tetris\_yesterday <- c(NA, tetris$tetris[1:25])  
run\_effect\_tetris\_yesterday <- lm(tetris$tetris\_yesterday ~ tetris$run)  
summary(run\_effect\_tetris\_yesterday)

##   
## Call:  
## lm(formula = tetris$tetris\_yesterday ~ tetris$run)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15610 -7521 -2337 2424 29220   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18903.8 3335.8 5.667 1.26e-05 \*\*\*  
## tetris$run 645.6 4823.5 0.134 0.895   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11560 on 21 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.0008524, Adjusted R-squared: -0.04673   
## F-statistic: 0.01792 on 1 and 21 DF, p-value: 0.8948

The effect of running on the tetris score yesterday was not zero, but 645.6212121 with a p-value of 0.895.

**d.) If Tetris responds to exercise, one might suppose that energy levels and GRE scores would as well. Are these hypotheses borne out by the data?**

run\_effect\_energy <- lm(energy ~ run, data = tetris)  
summary(run\_effect\_energy)

##   
## Call:  
## lm(formula = energy ~ run, data = tetris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.0714 -1.0179 0.0000 0.9286 1.9286   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.00000 0.33662 8.912 9.4e-09 \*\*\*  
## run 0.07143 0.44074 0.162 0.873   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.064 on 22 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.001192, Adjusted R-squared: -0.04421   
## F-statistic: 0.02627 on 1 and 22 DF, p-value: 0.8727

run\_effect\_gre <- lm(gre ~ run, data = tetris)  
summary(run\_effect\_gre)

##   
## Call:  
## lm(formula = gre ~ run, data = tetris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.8182 -0.6429 0.1818 0.3571 0.3571   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.8182 0.1385 5.909 5.05e-06 \*\*\*  
## run -0.1753 0.1850 -0.948 0.353   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4592 on 23 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.03757, Adjusted R-squared: -0.004275   
## F-statistic: 0.8978 on 1 and 23 DF, p-value: 0.3532

When regressed against energy, the run variable has a variable small effect size of 0.07143 with a p-value that is very high (0.873). Running effect on GRE question was actually negative, p-value of 0.353. Even though we might expect the effect of running on tetris to be similar to the effect of running on energy or gre, the data does not show the same causal effects.

**e.) Note that the observations in this regression are not necessarily all independent of each other. Given that, would you expect randomization inference to give you a better answer than the regression answer you just obtained in (b)? Which number(s) do you expect to be different in regression than in randomization inference? What is the direction of the bias?**

Randomization ensures that variables are independent of each other. Randomization inference could give a better answer if the study design was conducted to truly assign the treatments randomly. I would expect the effect on tetris score to be zero, but the effect on energy to be positive. The direction of bias is overestimating due to not discounting the increase in skill from playing the same game, testing a similar difficulty math question each day.