Problem Set 2

Research Design for Causal Inference
Due: April 14, 2015

Part I – Concepts

Complete the following exercises from FEDAI:

Chapter 2, Exercise 8:

(a) It appears that the intervention produced no effect on the proportion of applicants who had their residence verified. However, the applicants in the "Bribe" condition experienced a more than two-fold increase in the median speed with which their residence was verified (37 days in all other conditions versus 17 days in the "Bribe" condition).

(b) Among the three treatments, the "Bribe" and "RTIA" conditions had very large, positive effects on the proportion of applicants who received their ration cards within one year. As compared with the Control (20%) and "NGO" conditions (12.5%), all (100%) of the applicants in the "Bribe" condition and 83% of the applicants in the "RTIA" condition received their cards within a year.

(c) These results suggest that the Right to Information Act provides an extremely large improvement in the ability of slum dwellers who are unable or unwilling to provide a bribe to eventually obtain a ration card. However, the findings suggest that bribery is still the most reliable and efficient strategy. It is possible that some of the difference between the outcomes experienced by ration card applicants in the "Bribe" and "RTIA" conditions may be due to the applicants' widespread belief that bribery is the most effective method of acquiring a ration card. The research design cannot differentiate between the proportion of the differences due to this mechanism versus the other mechanisms related to the increased transparency enforced by the RTIA paperwork procedure.

Chapter 2, Exercise 12:

(a) The "natural" assignment of d_i likely has to do with specific prisoners' personal preference for reading over other kinds of activities. The distribution of this personal preference within the prisoner population may be systematically related to the distribution of prisoners' tendency toward violent encounters with prison staff. This might not only contribute to the systematically different rate of violence among those individuals actually observed in the "reading" versus "non-reading" conditions, but could also alter the counterfactual outcomes of the same individuals had they been assigned to the other condition (i.e., the formal expectations expressed in the question). It does not make sense to assume that reading (or not) would result in equal outcomes among the two sub-populations of prisoners.

(b) The answer to this question really depends on exactly what the original researcher's hypothesis was. If the original researcher's hypothesis is that reading for at least three hours per day reduces the likelihood of inmates' violent encounters with prison staff, then this design cannot meet the excludability assumption because the outcomes may be due to the reading, the solitary time in the prison library, or some other aspect of the treatment condition (maybe there are no staff in the library and thus fewer violent encounters!). If the hypothesis is that reading in specially designated carrels in the prison library for three hours causes reduced likelihood of violent encounters with prison staff, then the excludability assumption makes more sense.

(c) The non-interference assumption in this experiment entails that the potential outcomes for each prisoner reflect only the treatment or control status assigned to that prisoner and not the status of any other prisoner.

(d) The individual and overall effects of the program may vary depending on the proportion of prisoners assigned to the reading program.

Part II – Application

In this section, you will calculate descriptive statistics and estimate treatment effects for a subset of the data from a very famous experiment, Project STAR (Student-Teacher Achievement Ratio). As part of the Project STAR study, teachers, kindergarten students, and schools in Tennessee were randomly assigned into classrooms of varying sizes in order to estimate the effect of classroom size on student achievement. This study has launched dozens of papers, and you can read a brief summary of it on pp. 36-37 of the Willett & Murnane book. For a similar analysis of a different experiment, you may want to review Table 4.1 of Willett & Murnane on p. 49 and the surrounding discussion.

We'll be looking at a simplified cross-section of the data from 1985-1989 and will only compare the effects of one of the treatment conditions – a small classroom with 13-17 students – against the control condition – a regular classroom with 22-25 students – on one of the outcomes – reading test scores. For the purpose of this assignment do not worry about either the school-level aspects of the randomization/analysis or the additional control condition in the study. **In other words, pretend that random assignment occurred at the individual level and exclusively between the small and regular classroom size conditions.**

You can download the dataset for this assignment from:

http://aaronshaw.org/teaching/2015/causal/data/star.csv

The units of analysis (rows in the dataset) are individual students. The variables are listed in Table 1:

Table 1: Variables in simplified STAR experiment dataset

Variable name	Definition
`class.size`	Indicator of the student's class size ("small" or "regular").
`free.lunch`	Does the student receive free lunch or not?
`race`	The student's race (coded "black," "white," or "other").
`read.score`	The student's reading test score.
`gender`	The student's gender (coded either "male" or "female").
`teach.exper`	The number of years experience of the student's teacher.
`id`	A unique numeric identifier for each subject.

Question 1 – descriptive statistics

Report summary statistics for all of the pre-treatment covariates – both for the whole dataset and for the treatment and control groups respectively. For the continuous variable `teach.exper`, include minimum, mean, maximum, and standard deviation. For the categorical variables (`gender`, `free.lunch`, and `race`), present the number of subjects in each category.

```r
# First, download and read the dataset into a data frame:
d <- read.csv("http://aaronshaw.org/teaching/2015/causal/data/star.csv")

# Check to make sure the variables loaded correctly.
#
# Visually inspect the first few rows of the data set:
head(d)
```

```
##   read.score class.size teach.exper gender free.lunch  race id
## 1        447      small           7 female         no white  2
## 2        450      small          21 female         no black  3
## 3        448    regular          16   male         no white 11
## 4        447      small           5   male        yes white 12
## 5        431    regular           8   male        yes white 13
## 6        451    regular           3 female         no white 21

# And use lapply() to see if the variables are the right class (e.g.,
# are numeric variables numeric?):
lapply(d, class)

## $read.score
## [1] "integer"
##
## $class.size
## [1] "factor"
##
## $teach.exper
## [1] "integer"
##
## $gender
## [1] "factor"
##
## $free.lunch
## [1] "factor"
##
## $race
## [1] "factor"
##
## $id
## [1] "integer"
```

```
# Now generate descriptive statistics #
summary(d$teach.exper)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   4.000   8.000   9.038  13.000  27.000

sd(d$teach.exper)
```

```
## [1] 5.726875

# You can use table() to describe categorical variables:
table(d$gender)

##
## female   male
##   1814   1919

table(d$free.lunch)

##
##   no  yes
## 1964 1769

table(d$race)

##
## black other white
##  1176    20  2537

#
# And now I’ll compare across the treatment & control conditions
d$treat <- d$class.size == "small" # This makes the subsets tidy

summary(d$teach.exper[d$treat])

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   4.000   8.000   8.991  13.000  27.000

summary(d$teach.exper[!d$treat])

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   4.000   9.000   9.078  13.000  24.000

sd(d$teach.exper[d$treat])

## [1] 5.730971

sd(d$teach.exper[!d$treat])
```

```
## [1] 5.724448

# for the categorical covariates:
table(d$gender, d$treat)

##
##          FALSE TRUE
##   female   972  842
##   male    1028  891

table(d$free.lunch, d$treat)

##
##       FALSE TRUE
##   no   1051  913
##   yes   949  820

table(d$race, d$treat)

##
##         FALSE TRUE
##   black   636  540
##   other    10   10
##   white  1354 1183
```

Question 2 – assess covariate balance

Use t-tests and χ^2 tests to assess whether the treatment and control groups are "balanced" on the observed covariates. Summarize the results of these tests in a couple of sentences.

```
t.test(d$teach.exper[d$treat], d$teach.exper[!d$treat])

##
##  Welch Two Sample t-test
##
## data:  d$teach.exper[d$treat] and d$teach.exper[!d$treat]
## t = -0.4641, df = 3654.596, p-value = 0.6426
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4557902  0.2813251
## sample estimates:
## mean of x mean of y
##  8.990767  9.078000

# Both of the following methods work:
summary(table(d$gender, d$treat))

## Number of cases in table: 3733
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 7.01e-05, df = 1, p-value = 0.9933

chisq.test(table(d$gender, d$treat))

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(d$gender, d$treat)
## X-squared = 0, df = 1, p-value = 1

chisq.test(table(d$free.lunch, d$treat))

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(d$free.lunch, d$treat)
## X-squared = 0.0023, df = 1, p-value = 0.9614

chisq.test(table(d$race, d$treat))

##
##  Pearson's Chi-squared test
##
## data:  table(d$race, d$treat)
## X-squared = 0.2669, df = 2, p-value = 0.8751
```

Summary: These comparisons suggest no significant differences in the distribution of any pre-treatment covariates across the treatment and control groups.

Question 3 – estimate treatment effects

Estimate the Average Treatment Effect (ATE) using the average difference in the outcome variable (difference-in-means) as your estimator. Formally (using notation preferred by Willett & Murnane), I want you to calculate $\hat{\tau}$ when:

$$\hat{\tau} = \bar{y}_t - \bar{y}_c \tag{1}$$

```
# Compare the effect of treatment on the outcome -- in this case,
# student reading scores:

tau.hat <- mean(d$read.score[d$treat], na.rm=TRUE) -
mean(d$read.score[!d$treat], na.rm=TRUE)

tau.hat

## [1] 5.899496
```

Question 4 – interpretation

What do you conclude about the effect of this intervention based on these analyses?

> **Interpretation:** Based on these results, small class size increased average student reading test scores by almost 6 points.

Question 5

Why would you compare pre-treatment covariates (like I asked you to do in Question 2)? What do you learn from such comparisons?

> Comparing the distribution of pre-treatment covariates helps assess whether the subjects of the study assigned to treatment and control were "equal in expectation" before the treatment was administered. Formally, under the null hypothesis of no treatment effects:
>
> $$E[Y_i(1)] = E[Y_i(0)] \tag{2}$$
>
> For this expectation to be credible, it is important that the treatment status of any given unit in the experiment is independent of all covariates. Formally, random assignment ensures:
>
> $$(D_i \perp\!\!\!\perp Y_i) \mid \mathbb{X}_i \tag{3}$$

> That is, in the long run on average, our expectation of equality under the null hypothesis should hold up just fine because treatment assignment is indpendent of outcomes conditional on all covariates. However, just because we can expect this to be true doesn't ensure that our actually existing random treatment assignment eliminated all possible imbalance!
>
> By conducting the comparisons in question 2, we learn that the groups are, in fact, balanced for all observed pre-treatment covariates, suggesting that the equality of expectation under the null hypothesis is reasonable. We also learn that the random assignment procedure appears to have worked out just as probability theory suggests it should (whew!).
>
> You do not include any post-treatment measures in this comparison because such measures may have been affected by the treatment.

Question 6

Now, relax the assumption that random assignment occurred at the individual level. In other words, some of the individuals assigned to treatment may have been in the same classroom and/or school together (and likewise some of the individuals assigned to control). How and why might this change your interpretation of the estimate from Question 3?

> If the aforementioned assumptions do not hold up, then the outcomes (reading scores) for some students may be systematically related to the outcomes for other students due to causes unrelated to the treatment. In other words, students in a given classroom or school might perform better or worse on reading scores because of something to do with the particular cohort of students in the room or the behavior of the school principal, rather than the class size (treatment). These would all constitute violations of the assumption of non-interference – the treatment may have assumed different values for different students depending on the conditions under which it was administered.
>
> So, if the assumptions don't hold, I would feel much less certain about the naive estimate of average treatment effects calculated in Question 3 and would want to pursue a different estimation procedure that takes these aspects of the experimental conditions into account.

Part III – Key Concepts

Make sure to focus on understanding the following concepts as you read Gerber & Green, FEDAI, Chapter 3 this week:

- Sampling (or randomization) distribution.

- Standard deviation.
- Standard error.
- Variance.
- Covariance.
- P-value.
- Null hypothesis of no (average) effect.
- Randomization inference.
- Confidence intervals.
- Block random assignment.
- Cluster random assignment.