

# Section 6: review of fundamentals of causal inference (for experiments)

*Advanced Quantitative Methods (PLSC 504)*

*Fall 2017*

This handout reviews some of the fundamental concepts in causal inference (with experiments) that we have been discussing this semester with illustrations and accompanying R code.

## The setup:

Suppose we have a finite population of  $N$  units. To fix ideas, let  $N$  be the number of students at a particular high school in New Haven, CT. As in the last two homeworks, we are interested in studying whether taking a civics class affects “support for democracy”. Suppose that we have a legitimate way to measure this outcome and there is no measurement error involved.

The Neyman-Rubin Causal Model (NRCM) assumes each student  $i$  has two potential outcomes:  $Y_i(0)$  is their support for democracy without taking the civics class and  $Y_i(1)$  is their support for democracy if they do take the civics class. These are **fixed**.  $Y_i(0)$  is **not** a random variable. Neither is  $Y_i(1)$ . Fixed potential outcomes is an assumption, and we can argue about whether it is a reasonable one or not, but it doesn’t matter for our purposes so it’s safe to not worry about it and focus attention elsewhere.

The implication here is that each student has an individual treatment effect:  $\tau_i = Y_i(1) - Y_i(0)$ . This is their support for democracy under treatment (civics class taken) minus their support for democracy under control (civics class not taken). In words,  $\tau_i$  is the “effect of taking civics on student  $i$ ’s support for democracy”. We cannot observe this. This is the fundamental problem of causal inference.

We assume there is some underlying process that connects these potential outcomes to observed outcomes:  $Y_i = Y_i(1)W_i + Y_i(0)(1 - W_i)$ . This implies 1) there are no unobserved multiple version of treatment (consistency); 2) there is no interference between units – if person  $i$  gets treatment, this does not affect person  $j$ ’s potential outcome.

For the sake of this example, we will also assume that  $\tau_i = \tau \forall i$ . That is, the treatment effect is constant. This is a common assumption that is made. It’s almost certainly wrong in most social science applications, but still useful.<sup>1</sup>

To summarize:

1. We assume SUTVA:  $Y_i = Y_i(1)W_i + Y_i(0)(1 - W_i)$ .
2. The joint distribution of potential outcomes  $Y_i(0), Y_i(1)$  is unobserved. We just get to see  $(Y_i, W_i)$
3. The causal effect of treatment for unit  $i$  is  $\tau_i = Y_i(1) - Y_i(0)$ . We don’t get to see this. It’s unobserved. It’s fixed. We typically assume it’s constant.

Let’s set this up numerically to fix ideas. This is a toy example. Nature makes all these decisions for us. The researcher has no control over this process.

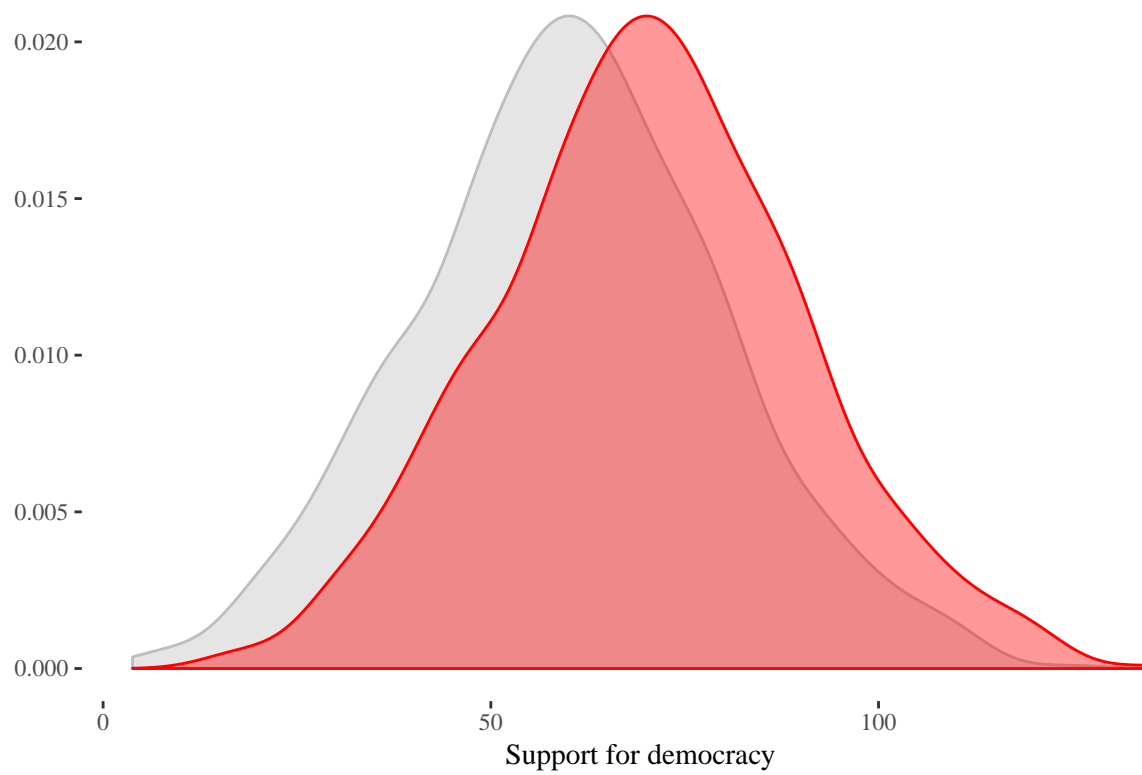
```
# A finite population example
N <- 1000
```

---

<sup>1</sup>Even if  $\text{Var}(\tau_i) \neq 0$  the difference in means estimator is still unbiased for the ATE. This is also true if we allow our potential outcomes to be stochastic, provided some mild regularity conditions hold. Want to break this? Assume  $Y_i(0) \sim \text{Cauchy}(\mu, \sigma)$  and even allow for a constant treatment effect. Pick any value for  $\mu, \sigma$ .

```
# Nature decides what  $Y_i(0)$ ,  $Y_i(1)$  are. For convenience, nature just  
# decides that  $Y_0$  is normally distributed.  
Y0 <- rnorm(n = N, mean = 60, sd = 20)  
  
# Nature also decided that the constant unit level treatment effect is 10  
Y1 <- Y0 + 10
```

Distribution of (unobserved) potential outcomes:



## Finite population causal inference

Suppose we have the budget to run an experiment on all 1000 students in the high school and our assignment mechanism is simple random assignment where  $\Pr(W_i = 1) = 1/2 \forall i$ . The estimand we want to learn about here is the average treatment effect (ATE):  $\frac{1}{N} \sum_i Y_i(1) - Y_i(0)$ . This quantity is just a function of the (unobserved) potential outcomes. It is fixed. Nature has already decided what it is,

```
tau <- Y1-Y0
ate <- mean(tau)
ate
```

```
## [1] 10
```

Think about what has happened above. Does the researcher have any control over this quantity? Can the researcher ever observe this? No. However, all is not lost. We know that the difference in means

$$\hat{\tau}_{DM} = \frac{1}{m} \sum_{i=1}^N Y_i W_i - \frac{1}{N-m} \sum_{i=1}^N Y_i (1 - W_i)$$

is an unbiased estimator for the ATE under this assignment mechanism, we have seen this proof many times. This is equivalent to writing,

$$\hat{\tau}_{DM} = \frac{\sum_{i=1}^N Y_i W_i}{\sum_i W_i} - \frac{\sum_{i=1}^N Y_i (1 - W_i)}{\sum_i (1 - W_i)}$$

Note that this estimator only depends on the observed data  $(Y_i, W_i)$ .  $Y_i(0)$  and  $Y_i(1)$  are **not** part of the observed data. This is an important point. This estimator,

$$\frac{1}{m} \sum_{i=1}^N Y_i(1) W_i - \frac{1}{N-m} \sum_{i=1}^N Y_i(0) (1 - W_i)$$

is undefined. Potential outcomes notation is reserved for estimands. This may seem pedantic. After all, we do see  $Y_i(1)$  when  $W_i = 1$ . Sure, but we don't see  $Y_i(1)W_i$  or  $Y_i(1)W_i(1)$ . It's an important distinction because it allows us to talk about identification and estimation with clarity.

So when does the researcher actually come onto the scene? Well, the researcher decides what their quantity of interest or estimand is. In this scenario it's the ATE. Then they choose an estimator. Here it's  $\hat{\tau}_{DM}$ . Identification and unbiased estimation of the ATE follows from the fact that the researcher is going to use the assignment mechanism already described (simple random assignment). The researcher has complete control over the assignment mechanism here,

```
# Assignment Mechanism. Decides what potential outcomes we observe!
W <- rbinom(n = N, size = 1, prob = 0.5)
```

This is the process that generates the observed outcome. We do not get to observe this process that connects unobserved potential outcomes to observed data. Under SUTVA,

```
Y <- Y1*W + Y0*(1-W)
```

What happened here? We randomly revealed the potential outcomes according to the realized  $W_i$ , the randomly assigned treatment. What's random here?  $W_i$  is obviously a random variable, we observe a particular realization of it.  $Y_i$  is also a random variable. Why? Because it is a function of the random variable  $W_i$ , and functions of random variables are random variables.  $Y_i(0)$  and  $Y_i(1)$  are fixed. Why? Because we assume that potential outcomes are fixed.

The estimator that we are using here is also a random variable. Why? Because it can be written as a function of random variables  $Y_i$  and  $W_i$ ,

```
# ATE (the estimator)
ate_hat <- sum(Y*W)/sum(W) - sum(Y*(1-W))/sum(1-W)
ate_hat
```

```
## [1] 11.28876
```

Uh oh. This isn't equal to the true ATE. We know the true ATE is 10 because we're playing god here. What happen? This is a single estimate. Unbiasedness is a property of an **estimator**, and this estimator is indeed unbiased,

```
nsims <- 100000

ate_hat_sim <- rep(NA, nsims)

for(i in 1:nsims) {

  W <- rbinom(n = N, size = 1, prob = 0.5)
  Y <- Y1*W + Y0*(1-W)
  ate_hat_sim[i] <- sum(Y*W)/sum(W) - sum(Y*(1-W))/sum(1-W)

}

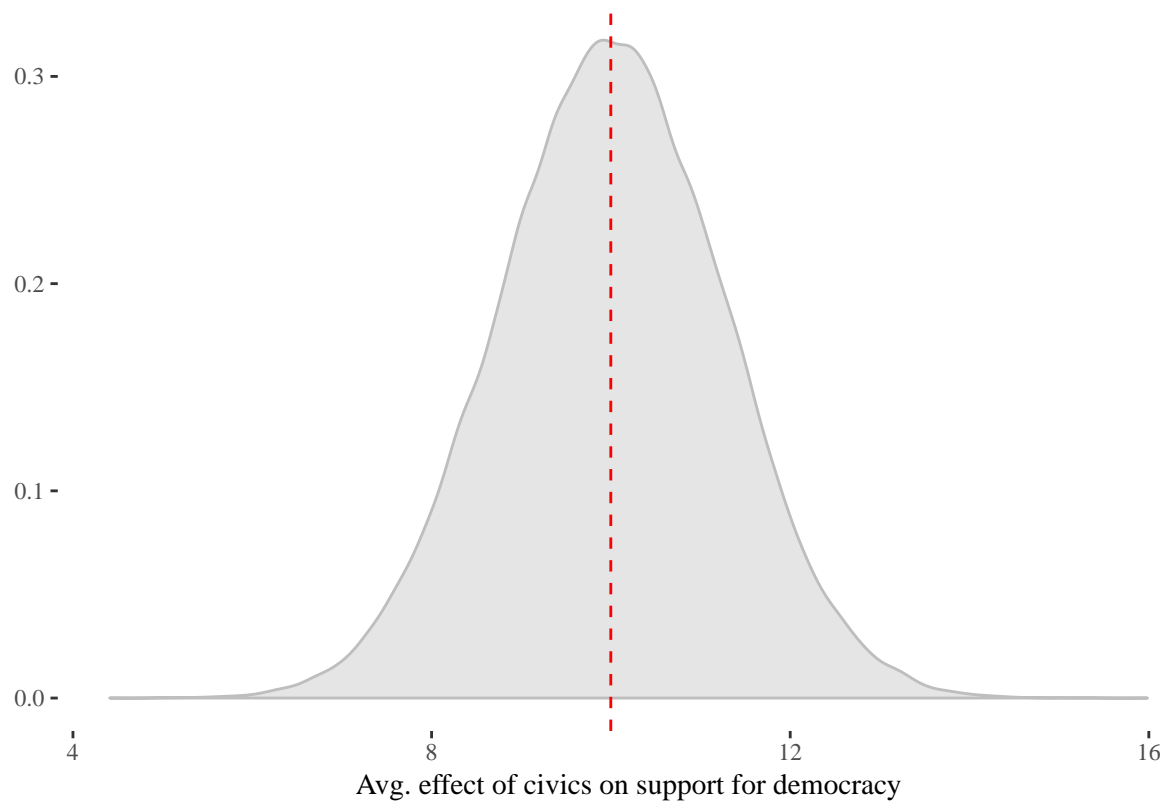
# Unbiased.
mean(ate_hat_sim)
```

```
## [1] 9.998674
```

```
# What's the variance (and MSE) of this estimator?
var(ate_hat_sim)
```

```
## [1] 1.568315
```

### Distribution of difference in means estimator:



Note that in this simulation to assess bias, the potential outcomes are fixed over the 100,000 estimates we obtain. The only source of randomness is  $W_i$  (and  $Y_i$  by implication).

## Causal inference: sampling from a finite population

Now we will introduce sampling uncertainty into the design. Suppose that we only have the budget to run the experiment on 200 students. The setup is otherwise the same.

Under this new research design  $S_i$  is 1 if the student is selected to be in the experimental sample, and 0 otherwise.  $W_i$  is still the binary treatment assignment.  $Y_i$  is still the observed outcome.  $Y_i(0)$  and  $Y_i(1)$  are still the unobserved potential outcomes. Under this design  $(S_i, W_i, Y_i)$  are random;  $Y_i(0)$  and  $Y_i(1)$  are still fixed.

In this scenario, we still wish to make inferences about the ATE in the *finite population* of 1000 students. Since we are now introducing sampling, we call this estimand the “population average treatment effect” or PATE,

$$\tau_{\text{PATE}} = \frac{1}{N} \sum_{i=1}^N Y_i(1) - Y_i(0) = \frac{1}{N} \sum_{i=1}^N \tau_i$$

Another estimand is the “sample average treatment effect” or SATE,

$$\tau_{\text{SATE}} = \frac{S_i [Y_i(1) - Y_i(0)]}{\sum_{i=1}^N S_i} = \frac{\sum_{i=1}^N S_i \tau_i}{\sum_{i=1}^N S_i}$$

Technically this estimand *is* a random variable because it depends on  $S_i$ . It’s not just a function of the fixed potential outcomes. The difference in means estimator (from before) is related via,

$$\hat{\tau}_{DM} = \tau_{\text{PATE}} + \epsilon_S$$

Where  $\epsilon_S$  is the “sampling error”. In the finite population example  $\epsilon_S = 0$ . Why?  $S_i = 1 \forall i$  so  $\sum_{i=1}^N S_i = N$ . Note also that  $\tau_{\text{PATE}} = \tau_{\text{SATE}}$  when  $\tau_i = \tau$ . Why? Because the PATE is a weighted average of the ATE for the units sampled (the SATE) plus the ATE for the units that are not sampled:

$$\frac{1}{N} \sum_{i=1}^N \tau_i = \frac{1}{2} \left( \frac{\sum_{i=1}^N S_i \tau_i}{\sum_{i=1}^N S_i} + \frac{\sum_{i=1}^N (1 - S_i) \tau_i}{\sum_{i=1}^N (1 - S_i)} \right)$$

When  $\tau_i = \tau$  (i.e  $\text{Var}(\tau_i) = 0$ ) then the LHS and RHS of this equation are equivalent.

Let’s work through the logic in R. First, take a random sample of 200 students from the high school,

```
# We only have a budget to run the experiment on 200. We take a
# random sample from the population of 1000
S <- as.numeric(1:N %in% sample(seq(1:N), 200))
```

Now, look at the things that the researcher doesn’t actually observe,

```
# What's the SATE (estimand)? Do we observe this?
sate <- sum(S*tau)/sum(S)
sate
```

```
## [1] 10
```

```
# What's the PATE (estimand)?
pate <- sum(tau)/N
pate == (1/2)*(sate + sum((1-S)*tau)/sum(1-S))
```

```
## [1] TRUE
```

Now the researcher steps onto the scene and applies treatment via simple random assignment on the sampled units.

```
# Do a random assignment
W <- rbinom(n = sum(S), size = 1, prob = 0.5)

# Again, we do not get to observe this process. Why?
Y <- Y1*S*W + Y0*S*(1-W)

# Diff in means: our estimator for the SATE.
sate_hat <- sum(Y*W)/sum(W) - sum(Y*(1-W))/sum(1-W)
sate_hat
```

```
## [1] -0.5264801
```

Uh oh. The difference in means isn't 10. This looks really bad! Is our estimator biased for the PATE? For the SATE? As before, bias is a property of an estimator. The difference in means is indeed unbiased for the PATE and the SATE in this setting,

```
nsims <- 100000
sate_hat_sim <- rep(NA, nsims)

for(i in 1:nsims) {

  # Take a random sample from the school
  S <- as.numeric(1:N %in% sample(seq(1:N), 200))

  # Do a simple random assignment
  W <- rbinom(n = sum(S), size = 1, prob = 0.5)

  # Reveal the potential outcomes
  Y <- Y1*S*W + Y0*S*(1-W)

  # Apply the difference in means estimator from before.
  sate_hat_sim[i] <- sum(Y*W)/sum(W) - sum(Y*(1-W))/sum(1-W)

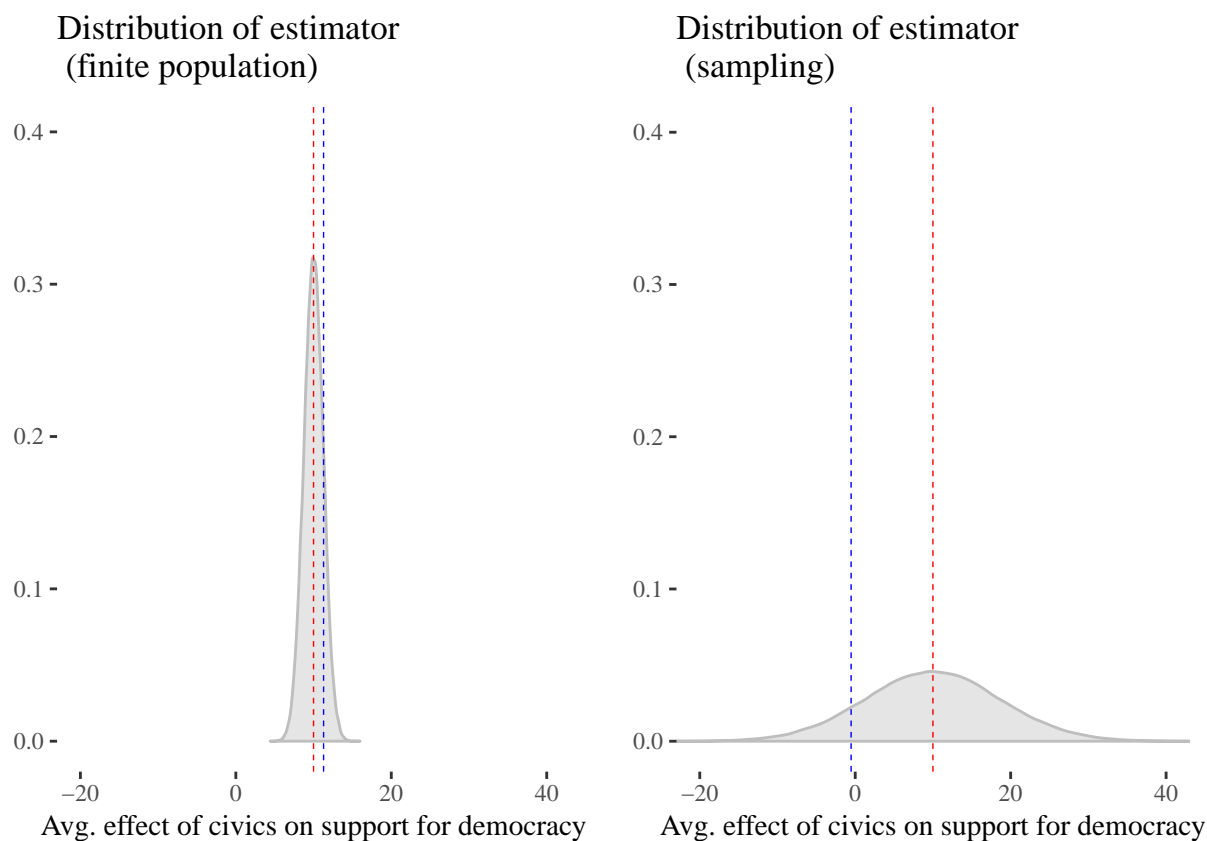
}

# Unbiased.
mean(sate_hat_sim)
```

```
## [1] 9.983088
```

```
# What's the variance (and MSE) of this *estimator*?
var(sate_hat_sim)
```

```
## [1] 75.98594
```



Note that the MSE of the difference in means estimator here is about 50 times higher (!) than the MSE of the difference in means estimator in the finite population example. What happened here? It's the same estimator as before!

Yes, but we have changed the **research design**. This new design adds sampling error into the mix. So how unusual is the estimate that we got under this new design? That is, what is the probability of observing something as extreme as the estimate we got?

```
sum(abs(sate_hat_sim - sate) >= abs(sate_hat - sate))/nsims
```

```
## [1] 0.22742
```

So it's not really that unusual. Another way to think about this. What's the “error distribution” (recall that  $\epsilon_s = \hat{\tau}_{DM} - \tau_{SATE}$  here)?



Distribution of sampling error

