# Randomization and the Randomization Distribution

## 1  Purpose

The purpose of this activity is to provide you with a thorough understanding of the statistical properties of randomization.

## 2  Overview

This lab can be completed during class time. We will work on it together as a class. The lab consists of two tasks. In each we will consider a hypothetical experiment, using each in turn to understand (1) the distinction between individual treatment effects and average treatment effects, and (2) the statistical properties of experimental designs.

## 3  Feedback and Assistance

Feedback and discussion of the activity will be provided during class. If you have questions, please come to my office hours.

## 4  Your Task

### The Omniscient Experimentalist

1. In the counterfactual interpretation of experiments originally described by Jerzy Neyman, Donald Rubin, and Paul Holland, an experiment randomly reveals "potential outcomes" for each unit, $i$, giving the researcher the value of one *manifest* outcome, $Y_i$ associated with the treatment that was randomly assigned and one or more unknown *potential* outcomes, $Y_i$, associated with the other experimental treatments. If an individual is assigned to condition 1, we therefore see manifest outcome $Y_{i1}$ but do not see potential outcomes $Y_{i2}$, $Y_{i3}$, etc. $\forall X$. We sometimes refer to this set of values as the "schedule of potential outcomes" for each unit.

   Table 1 describes a schedule of potential outcomes for a hypothetical experiment in which the researcher desired to know whether providing microfinance loans *and* entrepreneurial training to poor individuals ($X = 1$) would increase personal income five years later ($Y$, measured in £) relative to loans alone ($X = 0$). The column displays the results of this experiment from the perspective of the "omniscient experimentalist" who is able to see both potential outcomes for each participant.

2. Given omniscience about the full schedule of potential outcomes for every unit, what is each unit's individual treatment effect, $TE_i$, where $TE_i = Y_{i1} - Y_{i0}$?

3. What is the average treatment effect, $ATE$? Is the training programme effective?

Table 1: Schedule of Potential Outcomes for Entrepreneurship Programme

| Unit | Treatment, $X$ | $Y_{i0}$ (£) | $Y_{i1}$ (£) | $TE_i$ (£) |
|------|------|------|------|------|
| 1 | 0 | 10,000 | 12,000 | |
| 2 | 0 | 12,000 | 13,000 | |
| 3 | 1 | 8,000 | 10,000 | |
| 4 | 1 | 9,000 | 15,000 | |
| 5 | 0 | 11,000 | 9,000 | |
| 6 | 1 | 5,000 | 20,000 | |
| 7 | 0 | 7,000 | 7,000 | |
| 8 | 1 | 8,000 | 11,000 | |
| 9 | 1 | 10,000 | 7,000 | |
| 10 | 0 | 13,000 | 14,000 | |
| 11 | 1 | 10,000 | 12,000 | |
| 12 | 1 | 4,000 | 8,000 | |
| 13 | 0 | 10,000 | 21,000 | |
| 14 | 0 | 13,000 | 9,000 | |
| 15 | 0 | 8,000 | 12,000 | |
| 16 | 1 | 9,000 | 9,000 | |
| ATE | – | | | |

4. If we could know the full schedule of potential outcomes, which of the following alternative policies would maximize welfare (i.e., maximize average income for all participants):

   (a) All units receive only loans

   (b) All units receive loans and training

   (c) All units receive whichever intervention maximizes their personal income

5. If the training programme costs £500 per participant, do you consider the programme to be *cost effective* given its impact?

## Randomization Distributions

6. In addition to knowing whether an intervention produces any difference in outcomes between experimental conditions, we also typically want to know whether that ATE is statistically distinguishable from no effect. In other words, we want to know whether that statistic — the average treatment effect — is *statistically significant*.

   To assess statistical significance, we need to construct a "sampling distribution" for the statistic that conveys the variation in estimates of an ATE that is actually 0 due to chance variation and see how unusual the ATE we found is relative to that null hypothesis sampling distribution.

   With experiments, we can construct a specific kind of sampling distribution called a "randomization distribution." The randomization distribution uses the data we observe from our experiment to create a set of possible values of the ATE that we would see in these data if the true ATE were 0 (i.e., there is no difference between the two treatments) and

we simply shuffled which units were assigned to which treatment. In the "Omniscient Experimentalist" example, there were 16 units of which 8 received treatment. This means there are $\binom{n}{n/2} = \binom{16}{8} = 12870$ possible randomization schemes. We cannot work with all of those by hand, so we'll use a simpler example for this exercise.

Consider a hypothetical experiment in which LSE wants to know whether investing in targeted recruitment activities at sixth-form schools improves the number of students that apply to LSE from those schools. To assess this, the Student Recruitment Office conduct an experiment in which eight schools of similar size and other characteristics participate. Four school are randomly assigned to receive targeted recruitment ($X = 1$) and the remaining four are randomly assigned to receive no recruitment activities ($X = 0$) and the percentage of students from each school applying to LSE the subsequent year is measured ($Y$).

The four schools assigned to control (labelled A, B, C, D) generated application rates of 5%, 7%, 9%, and 4% respectively. The four schools assigned to treatment (labelled E, F, G, H) generated application rates of 9%, 4%, 13%, and 12% respectively. What are the mean outcomes in the two groups and what is the $\widehat{ATE}$ for this experiment?

7. To assess the statistical significance of this effect (i.e., whether it is larger than would be expected by chance), we need to construct the randomization distribution for the ATE. Table 2 shows the complete $\binom{8}{4} = 70$ possible randomizations of the experiment. For each randomization, calculate $\bar{Y}_0$, $\bar{Y}_1$, and $\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0$. The distribution of $\widehat{ATE}$ is the randomization distribution. Calculate every possible ATE estimate that could have resulted from this experiment under alternative randomizations.

8. How likely would it be to see an estimated ATE of the size we observed in the experiment if the true effect size were actually 0?

9. Another way of using the randomization distribution is as a way of constructing a standard error for the ATE. The standard deviation of the randomization distribution is $SE(ATE)$. What is this standard error?

10. Bonus question: use the R to approximate the randomization distribution and assess the statistical significance of the results. Then, compare this to a simple t-test. (Solution code on the final page.)

Table 2: All Possible Randomizations

| Unit: | A | B | C | D | E | F | G | H | | $\bar{Y}_0$ | $\bar{Y}_1$ | $\widehat{ATE}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y_i$: | 5 | 7 | 9 | 4 | 9 | 4 | 13 | 12 | | | | |
| Randomization | | | | | | | | | | | | |
| 1 (Actual) | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | | | | |
| 2 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | | | | |
| 3 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | | | | |
| 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | | | | |
| 5 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | | | | |
| 6 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | | | | |
| 7 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | | | | |
| 8 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | | | | |
| 9 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | | | | |
| 10 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | | | | |
| 11 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | | | | |
| 12 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | | | | |
| 13 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | | | | |
| 14 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | | | | |
| 15 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | | | | |
| 16 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | | | | |
| 17 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | | | | |
| 18 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | | | | |
| 19 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | | | | |
| 20 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | | | | |
| 21 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | | | | |
| 22 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | | | | |
| 23 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | | | | |
| 24 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | | | | |
| 25 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | | | | |
| 26 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | | | | |
| 27 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | | | | |
| 28 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | | | | |
| 29 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | | | | |
| 30 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | | | | |
| 31 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | | | | |
| 32 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | | | | |
| 33 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | | | | |
| 34 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | | | | |
| 35 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | | | | |

| Unit: | A | B | C | D | E | F | G | H | | $\bar{Y}_0$ | $\bar{Y}_1$ | $\widehat{ATE}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y_i$: | 5 | 7 | 9 | 4 | 9 | 4 | 13 | 12 | | | | |
| Randomization | | | | | | | | | | $\bar{Y}_0$ | $\bar{Y}_1$ | $\widehat{ATE}$ |
| 36 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | | | | |
| 37 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | | | | |
| 38 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | | | | |
| 39 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | | | | |
| 40 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | | | | |
| 41 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | | | | |
| 42 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | | | | |
| 43 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | | | | |
| 44 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | | | | |
| 45 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | | | | |
| 46 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | | | | |
| 47 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | | | | |
| 48 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | | | | |
| 49 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | | | | |
| 50 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | | | | |
| 51 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | | | | |
| 52 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | | | | |
| 53 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | | | | |
| 54 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | | | | |
| 55 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | | | | |
| 56 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | | | | |
| 57 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | | | | |
| 58 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | | | | |
| 59 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | | | | |
| 60 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | | | | |
| 61 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | | | | |
| 62 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | | | | |
| 63 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | | | | |
| 64 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | | | | |
| 65 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | | | | |
| 66 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | | | | |
| 67 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | | | | |
| 68 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | | | | |
| 69 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | | | | |
| 70 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | | | | |

# Code for randomization inference in R

```
# construct data
d <- data.frame(x = c(0,0,0,0,1,1,1,1), y = c(5,7,9,4,9,4,13,12))

# calculate ATE from each randomization
set.seed(1)     # set random number seed
n <- 10000      # number of randomizations
rd <- replicate(n, coef(lm(d$y ~ sample(d$x, 8)))[2])

# visualize the randomization distribution
hist(rd)
abline(v = coef(lm(y~x, data = d))[2L])

# one-tailed significance test
sum(rd >= coef(lm(y ~ x, data = d))[2L])/n
# two-tailed significance test
sum(abs(rd) >= coef(lm(y ~ x, data = d))[2L])/n

# one-tailed t-test
t.test(y ~ x, data = d, alternative = "less")
# two-tailed t-test
t.test(y ~ x, data = d, alternative = "two.sided")
```