2. ESTIMATING CAUSAL EFFECTS WITH RANDOMIZED EXPERIMENTS

One of the main purposes of data analysis in the social sciences is the estimation of causal effects, also known as causal inference. What are causal effects? And how can we best estimate them? These are the main questions we answer in this chapter. To illustrate the concepts covered, we analyze data from a real-world experiment. Specifically, we estimate the causal effect of small classes on student performance using data from Project STAR.

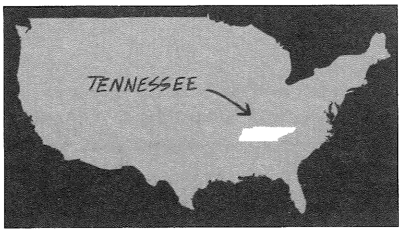
R symbols, operators, and functions introduced in this chapter: ==, ifelse(), and [].

2.1 PROJECT STAR

In the 1980s, Tennessee legislators began to consider reducing class size in the state's schools in an effort to improve student performance. Some studies had suggested that smaller classes are more conducive to learning than regular-size classes, especially in the early schooling years. Reducing class size, however, would require additional funds to pay for the extra teachers and classroom space. Before moving forward with the new policy, the legislature decided to commission a thorough investigation of the causal effects of small classes on student performance. The result was a multimillion-dollar study called Project Student-Teacher Achievement Ratio (Project STAR).

Based on Frederick Mosteller, "The Tennessee Study of Class Size in the Early School Grades," *Future of Children* 5, no. 2 (1995): 113–27. We study the effects of small classes as compared to regular-size classes (without aides), disregarding data from students who were assigned to regular-size classes with aides. We focus on the initial group of participants who were randomly assigned to different class types before entering kindergarten and exclude observations with any missing data in the variables used in the analysis.

In this chapter, we analyze a portion of the data from Project STAR. The aim of the project was to examine the effects of class size on student performance in both the short and long term. The project consisted of an experiment in which kindergartners were randomly assigned to attend either small classes, with 13 to 17 students, or regular-size classes, with 22 to 25 students, until the end of third grade. Researchers followed student progress over time. As the outcome variables of interest, we have student scores on third-grade standardized tests in reading and math as well as high school graduation rates.



2.2 TREATMENT AND OUTCOME VARIABLES

Tennessee legislators wanted researchers to estimate the causal effects of small classes on educational outcomes. Specifically, they wanted to know whether student performance improves as a direct result of attending a small class and not just as a result of other factors that may accompany small class sizes, such as better teachers, higher-performing classmates, or greater resources.

A **causal relationship** refers to the cause-and-effect connection between:

- the **treatment variable** (X): variable whose change may produce a change in the outcome variable
- the **outcome variable** (Y): variable that may change as a result of a change in the treatment variable.

Causal relationships refer to the cause-and-effect connection between two variables. In this case, the two variables are (i) small class and (ii) student performance.

In this book, we study causal relationships in which there is clear directionality in how the two variables relate to each other: changes in one variable may cause changes in the other. We use this directionality to distinguish between the variables. We refer to the variable where the change originates as the **treatment variable**. We refer to the variable that may change in response to the change in the treatment variable as the **outcome variable**. Here, small class is the treatment variable, and student performance is the outcome variable.

TIP: At some point, you might have learned about dependent and independent variables. Treatment variables are a type of independent variable, and outcome variables are the same as dependent variables.

In mathematical notation, we represent the treatment variable as X and the outcome variable as Y. We represent the causal relationship between them visually with an arrow from X to Y. The direction of the arrow indicates that changes in X may produce changes in Y but not the other way around:

$$X \rightarrow Y$$

In Project STAR, we are interested in the following causal link:

$$\textit{small class} \;\rightarrow\; \textit{student performance}$$

The distinction between treatment and outcome variables depends on the nature of the causal relationship between them as well as on the research question. The same variable might be the outcome in one study but be the treatment in another. For example, in one study we may be interested in the effect of attending a small class on the probability of graduating from high school. Here, the variable that indicates whether a student graduated from high school, *graduated*, is the outcome variable (diagram A below). In another study, we may be interested in the effect of graduating from high school on future wages. In that case, *graduated* would be the treatment variable (diagram B).

(A) $\textit{small class} \rightarrow \textit{graduated}$
(B) $\textit{graduated} \rightarrow \textit{future wages}$

2.2.1 TREATMENT VARIABLES

In this book, for the sake of simplicity, we focus on treatment variables that are binary, that is, that indicate whether the treatment is present or absent. We define the treatment variable for each individual i as:

RECALL: A binary variable takes only two values, in this book 1s and 0s, and the notation i identifies the position of the observation in a dataframe or in a variable.

$$X_i = \begin{cases} 1 \text{ if individual } i \text{ receives the treatment} \\ 0 \text{ if individual } i \text{ does not receive the treatment} \end{cases}$$

Based on whether the individual receives the treatment, we speak of two different conditions:

- treatment is the condition with the treatment (X_i=1)
- control is the condition without the treatment (X_i=0).

Two conditions:
- **treatment**: when X_i=1
- **control**: when X_i=0.

We describe the observations that receive the treatment as being *under treatment* or *treated* and those that do not as being *under control* or *untreated*.

For example, in the analysis of the STAR dataset, we are interested in examining the effects of attending a small class on student performance. The treatment variable, which we name *small*, is a binary variable that equals 1 if the student attended a small class and 0 otherwise. Formally, we define *small* as:

RECALL: In the STAR dataset, each observation i represents a different student because the unit of observation is students.

$$small_i = \begin{cases} 1 \text{ if student } i \text{ attended a small class} \\ 0 \text{ if student } i \text{ did not attend a small class} \end{cases}$$

2.2.2 OUTCOME VARIABLES

We will see different types of outcome variables. For example, in the STAR dataset, we will analyze the effect of attending a small class on three different measures of student performance: *reading*, *math*, and *graduated*. While the first two outcome variables are non-binary, the third is binary. As we will see later in the chapter, the interpretation of the results depends on the type of outcome variable used in the analysis.

2.3 INDIVIDUAL CAUSAL EFFECTS

When estimating the causal effect of X on Y, we attempt to quantify the change in the outcome variable Y that is caused by a change in the treatment variable X. For example, if interested in the effect of *small* on *reading*, we aim to measure the extent to which student performance on the reading test improves or worsens as a result of attending a small class, as opposed to a regular-size class.

The **causal effect of** X **on** Y is the change in the outcome variable Y caused by a change in the treatment variable X.

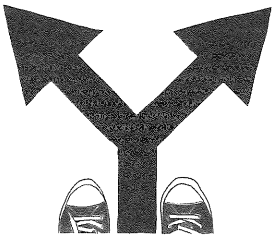
Note that when estimating a causal effect, we are trying to measure a *change* in Y, specifically the change in Y caused by a change in X. In mathematical notation, we represent change with $\triangle$ (the Greek letter Delta), and thus, we represent a change in the outcome as $\triangle Y$.

Two potential outcomes:

- **potential outcome under the treatment condition** (the value of Y_i if X_i=1)
- **potential outcome under the control condition** (the value of Y_i if X_i=0).

To measure this change in the outcome Y, ideally we would compare two potential outcomes: the outcome when the treatment is present and the outcome when the treatment is absent. In mathematical notation, we represent these two potential outcomes as follows:

- $Y_i(X_i{=}1)$ represents the **potential outcome under the treatment condition** for individual i (the value of Y_i if X_i=1)
- $Y_i(X_i{=}0)$ represents the **potential outcome under the control condition** for individual i (the value of Y_i if X_i=0).



If, for each individual i, we could observe both potential outcomes, then computing the change in the outcome Y caused by the treatment X would be simple. We would just compute the difference between these two potential outcomes. Mathematically, the individual causal effects of receiving the treatment X on the outcome Y would be computed as shown in formula 2.1.

FORMULA 2.1. Definition of the individual causal effects of a treatment on an outcome.

IF WE COULD OBSERVE
BOTH POTENTIAL OUTCOMES

$$individual\_effects_i = \triangle Y_i = Y_i(X_i{=}1) - Y_i(X_i{=}0)$$

where:

- $\triangle Y_i$ is the change in the outcome individual i would have experienced by receiving the treatment, as compared to not receiving the treatment
- $Y_i(X_i{=}1)$ and $Y_i(X_i{=}0)$ are the two potential outcomes for the same individual i, under the treatment and the control conditions, respectively.

For example, if we are estimating the effect of attending a small class on reading test scores using the data from Project STAR, the treatment variable X would be *small* and the outcome variable Y would be *reading*. In this case, for each student i, we would like to observe third-grade reading test scores both (i) after attending a small class from kindergarten to third grade and (ii) after attending a regular-size class from kindergarten to third grade. If this were possible, we could directly measure the causal effect that attending a small class had on each student's reading performance by calculating:

$$\triangle reading_i = reading_i(small_i{=}1) - reading_i(small_i{=}0)$$

where:

- $\triangle reading_i$ is the change in reading test scores student i would have experienced by attending a small class, as compared to a regular-size class
- $reading_i(small_i{=}1)$ is the third-grade reading test score of student i after attending a small class (the value of $reading_i$ if $small_i{=}1$)
- $reading_i(small_i{=}0)$ is the third-grade reading test score of the same student i after attending a regular-size class (the value of $reading_i$ if $small_i{=}0$).

TIP: This is formula 2.1 with *reading* as the Y variable and *small* as the X variable. If we could observe both potential outcomes for every student, we could use this formula to compute the individual causal effects of attending a small class on reading test scores.

Let's imagine, for a moment, that we *could* observe both potential outcomes for each of the first six students in the STAR dataset. See the first two columns of table 2.1 below. For illustration purposes, we made up the values of the potential outcomes that were not observed (shown in gray). If these were indeed the true potential outcomes, then the individual causal effects of *small* on *reading* for these six students would be the values shown in the third column of table 2.1.

i	$reading(small{=}1)$	$reading(small{=}0)$	$\triangle reading$
1	578	571	7
2	611	612	-1
3	586	583	3
4	661	661	0
5	614	602	12
6	607	610	-3

TABLE 2.1. If for each student i, we could observe both potential outcomes, then we could measure the causal effects of *small* on *reading* at the individual level. (Warning: Here we made up the values of the unobserved potential outcomes, shown in gray, for the sake of illustrating individual causal effects.)

Based on table 2.1, we would conclude that attending a small class as opposed to a regular-size one:

- increased the reading score of the first student by 7 points, the score of the third student by 3 points, and the score of the fifth student by 12 points
- decreased the reading score of the second student by 1 point, and the score of the sixth student by 3 points
- had no effect on the reading score of the fourth student.

TIP: When interpreting the sign of causal effects, we should interpret:

- a positive effect as the treatment causing an *increase* in the outcome variable
- a negative effect as the treatment causing a *decrease* in the outcome variable
- an effect of zero as the treatment causing *no change* in the outcome variable.

Notice that the same treatment might have different effects for different individuals. In addition, note that since a causal effect is a measure of change, we should interpret a causal effect as an increase if positive, as a decrease if negative, and as having no effect if zero. (See TIP in the margin.)

A different way of expressing the two potential outcomes:

- **the factual outcome**: potential outcome under whichever condition (treatment or control) was received in reality
- **the counterfactual outcome**: potential outcome under whichever condition (treatment or control) was not received in reality

Unfortunately, this kind of analysis is not possible. In the real world, we never observe both potential outcomes for the same individual. Instead, we observe only the **factual outcome**, which is the potential outcome under whichever condition (treatment or control) was received in reality. We can never observe the **counterfactual outcome**, which is the potential outcome that would have occurred under whichever condition (treatment or control) was not received in reality. As a result, we cannot compute causal effects at the individual level. In our example, a student attends either a small or a regular-size class during the early schooling years but cannot enter a parallel universe to attend both at the same time. (See figure 2.1.)

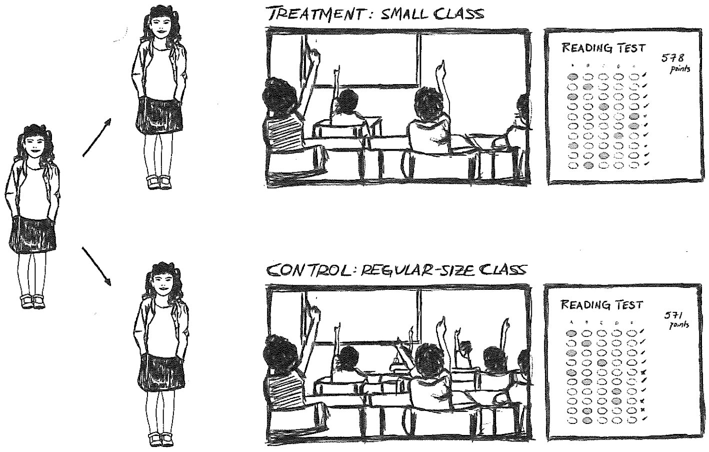


FIGURE 2.1. If an individual could split into two identical beings, and each one of them could receive a different condition, then we could observe the outcome under the treatment condition and the outcome under the control condition for the same individual. We could then calculate the causal effect of the treatment on the outcome for this specific individual by simply measuring the difference between the two outcomes.

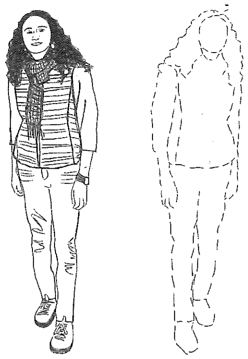
For each student in Project STAR, for instance, we observe only one third-grade reading test score, the score earned after the student actually attended one of the two types of classes. As a result, we cannot measure how class size affected each student's performance on the reading test. (See table 2.2, where the counterfactual outcomes for the first six observations are indicated as ??? because they were unobserved.)

TABLE 2.2. Values of *small*, *reading*, *reading*(*small*=1), and *reading*(*small*=0) for the first six observations in the STAR dataset. Unobserved potential outcomes, or counterfactuals, are indicated as ???.

i	*small*	*reading*	*reading*(*small*=1)	*reading*(*small*=0)
1	1	578	578	???
2	0	612	???	612
3	0	583	???	583
4	1	661	661	???
5	1	614	614	???
6	0	610	???	610

Take the first student, the observation when i=1. The value of $small_1$ is 1, which means this student attended a small class. The value of $reading_1$, then, indicates the performance of this student on the reading test after attending a small class ($reading_1(small_1{=}1){=}578$). The score of 578 points is this student's factual outcome because we did observe it. What we did not observe is the counterfactual outcome, that is, how this student would have performed on the reading test after attending a regular-size class ($reading_1(small_1{=}0){=}???$). Consequently, we cannot measure the effect attending a small class had on this student's reading test score:

$$\begin{aligned}\triangle reading_1 &= reading_1(small_1{=}1) - reading_1(small_1{=}0)\\ &= 578 - ??? = ???\end{aligned}$$



We observe only what happens in reality (the factual outcome). We can never observe what would have happened had we made different decisions (the counterfactual outcome).

The fundamental problem we face when inferring causal effects is that we never observe the same individual both receiving the treatment and not receiving the treatment at the same time.

FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE: To measure causal effects, we need to compare the factual outcome with the counterfactual outcome, but we can never observe the counterfactual outcome.

2.4 AVERAGE CAUSAL EFFECTS

To get around the fundamental problem of causal inference, we must find good approximations for the counterfactual outcomes. To accomplish this, we move away from individual-level effects and focus on the *average* causal effect *across a group of individuals.*

RECALL: The average of a variable equals the sum of the values across all observations divided by the number of observations. It is often represented by the name of the variable with a bar on top.

The average causal effect of the treatment X on the outcome Y, also known as the average treatment effect, is the average of all the individual causal effects of X on Y within a group. Since each individual causal effect is the change in Y caused by a change in X for a particular individual, the average causal effect of X on Y is the *average* change in Y caused by a change in X *for a group of individuals.*

The **average causal effect** of X on Y, also known as the **average treatment effect**, is defined as the average of the individual causal effects of X on Y across a group of individuals. It is the average change in Y caused by a change in X for a group of individuals.

If we could observe both potential outcomes for each individual in the group, then we could measure individual causal effects (using formula 2.1) and compute the average causal effect as shown in formula 2.2.

FORMULA 2.2. Definition of the average causal effect of a treatment on an outcome, or the average treatment effect.

IF WE COULD OBSERVE
BOTH POTENTIAL OUTCOMES

$$\overline{individual\_effects} = \frac{\sum_{i=1}^{n} individual\_effects_i}{n}$$

where:

- $\overline{individual\_effects}$ is the average causal effect for the observations in the study, and $individual\_effects_i$ is the individual causal effect for observation i
- $\sum_{i=1}^{n} individual\_effects_i$ stands for the sum of all $individual\_effects_i$ from i=1 to i=n, meaning from the first observation of $individual\_effects$ to the last one
- n is the number of observations in the study.

Let's return to the idealized scenario where we could observe both potential outcomes for each of the first six students in the STAR dataset. As we saw earlier, if the potential outcomes were those shown in table 2.1, the individual causal effects of *small* on *reading* for these students would be:

$$individual\_effects = \{7, \text{-}1, 3, 0, 12, \text{-}3\}$$

Then, the average causal effect of *small* on *reading* would be:

$$\begin{aligned}\overline{individual\_effects} &= \frac{\sum_{i=1}^{n} individual\_effects_i}{\text{number of students}} \\ &= \frac{7 + (\text{-}1) + 3 + 0 + 12 + (\text{-}3)}{6} = \frac{18}{6} = 3\end{aligned}$$

We would conclude that, among the first six students in Project STAR, attending a small class, as opposed to a regular-size one, improved student performance on the reading test by 3 points, on average. Remember, though, this kind of analysis is not possible because we never observe both potential outcomes for the same individual. Therefore, we are not going to be able to compute average causal effects directly, either.

How can we obtain good approximations for the counterfactual outcomes, which by definition cannot be observed? As we will see in detail soon, we must find or create a situation in which the treated observations and the untreated observations are similar with respect to all the variables that might affect the outcome other than the treatment variable itself. The best way to accomplish this is by conducting a randomized experiment.

2.4.1 RANDOMIZED EXPERIMENTS AND THE DIFFERENCE-IN-MEANS ESTIMATOR

In a **randomized experiment**, also known as a randomized controlled trial (RCT), researchers decide who receives the treatment based on a random process.

A **randomized experiment** is a type of study design in which treatment assignment is randomized.

For example, in Project STAR, researchers could have flipped a coin to decide whether a student would attend a small or a regular-size class. If the coin landed on heads, the student would be assigned to a small class. If tails, the student would be assigned to a regular-size class. (See figure 2.2.)

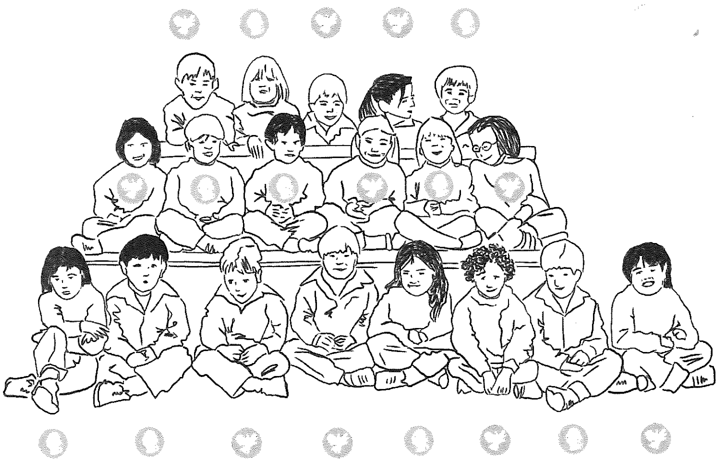
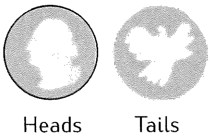


FIGURE 2.2. One way of assigning treatment at random is to flip a coin for every individual in the study. If the coin lands on heads, the individual is assigned to the treatment group. If tails, the individual is assigned to the control group.



In practice, researchers do not flip coins but instead use a computer program like R to assign at random a 1 or a 0 to each individual. Individuals who are assigned a 1 are given the treatment, and individuals who are assigned a 0 are not given the treatment.

Once the treatment is assigned, we can differentiate between two groups of observations:

- the **treatment group** consists of the individuals who received the treatment (the group of observations for which $X_i=1$)
- the **control group** consists of the individuals who did not receive the treatment (the group of observations for which $X_i=0$).

Two **groups**:

- **treatment group**: observations that received the treatment
- **control group**: observations that did not receive the treatment.

In Project STAR, the students who attended a small class are the treatment group. The students who attended a regular-size class are the control group.

When treatment assignment is randomized, the only thing that distinguishes the treatment group from the control group, besides the reception of the treatment, is chance. This means that although the treatment and control groups consist of different individuals, the two groups are comparable to each other, *on average*, in all respects other than whether or not they received the treatment.

Random treatment assignment makes the treatment and control groups *on average* identical to each other in all observed and unobserved pre-treatment characteristics. Pre-treatment characteristics are the characteristics of the individuals in a study before the treatment is administered. (By definition, pre-treatment characteristics cannot be affected by the treatment.)

Pre-treatment characteristics are the characteristics of the individuals in a study before the treatment is administered.

TIP: An unobserved characteristic is a characteristic that we have not measured.

For example, in Project STAR, since the treatment was randomly assigned, the average age of the treatment group—the students who attended a small class—should be comparable to the average age of the control group—the students who attended a regular-size class.

RANDOMIZATION OF TREATMENT ASSIGNMENT:
By randomly assigning treatment, we ensure that treatment and control groups are, on average, identical to each other in all observed and unobserved pre-treatment characteristics.

Let's return to the formula of the average treatment effect. If we could observe both potential outcomes for each individual, we could compute individual causal effects (using formula 2.1), and the average treatment effect would equal the average difference between the two potential outcomes:

$$\text{average\_effect} = \overline{\textit{individual\_effects}} = \overline{Y(X{=}1) - Y(X{=}0)}$$

By the rules of summation, the average of a difference is equal to the difference of averages. (For an example, see the TIP in the margin.) This allows us to rewrite the average treatment effect:

TIP: Using the values in the table below, we can confirm that the average of the difference between X and Y equals the difference between the average of X and the average of Y:

i	X	Y	$X-Y$
1	4	2	2
2	10	4	6
averages	7	3	4

$\overline{X-Y} = 4$ and $\overline{X}-\overline{Y} = 7-3 = 4$

$$\text{average\_effect} = \overline{Y(X{=}1) - Y(X{=}0)} = \overline{Y(X{=}1)} - \overline{Y(X{=}0)}$$

where:

- $\overline{Y(X{=}1)}$ is the average outcome under the treatment condition across all observations
- $\overline{Y(X{=}0)}$ is the average outcome under the control condition across all observations.

Unfortunately, we cannot compute the average treatment effect this way because, as you may recall, we never observe both potential outcomes for each individual. Therefore, we cannot compute either the average outcome under the treatment condition *across all observations* or the average outcome under the control condition *across all observations*. All we can observe is the average outcome for the treatment group after receiving the treatment and the average outcome for the control group after not receiving the treatment.

If the treatment and control groups were comparable before the treatment was administered, however, then we can use the factual outcome of one group as an approximation for the counterfactual outcome of the other. In other words, we can assume that the average outcome of the treatment group is a good estimate of the average outcome of the control group, had the control group received the treatment. Similarly, we can assume that the average outcome of the control group is a good estimate of the average outcome of the treatment group, had the treatment group not received the treatment. As a result, we can approximate the average treatment effect by computing the difference in the average outcomes between the treatment and control groups. Since both of these average outcomes *are* observed, this is an analysis we *are* able to perform.

To summarize, if the treatment and control groups were comparable before the treatment was administered, we can estimate the average causal effect of treatment X on outcome Y with formula 2.3, which is known as the **difference-in-means estimator**.

IF GROUPS WERE COMPARABLE BEFORE THE TREATMENT WAS ADMINISTERED

$$\widehat{\text{average\_effect}} = \overline{Y}_{\text{treatment group}} - \overline{Y}_{\text{control group}}$$

where:

- $\widehat{\text{average\_effect}}$ stands for the estimated average treatment effect (the "hat" on top of the name denotes that this is an estimate or approximation)
- $\overline{Y}_{\text{treatment group}}$ is the average outcome for the treatment group and $\overline{Y}_{\text{control group}}$ is the average outcome for the control group (both of which are observed).

FORMULA 2.3. The right-hand side of the equation is the formula for the **difference-in-means estimator**, which produces a valid estimate of the average treatment effect when the treatment and control groups are comparable with respect to all the variables that might affect the outcome other than the treatment variable itself.

TIP: To estimate causal effects, it is necessary to have both a treatment group and a control group. In other words, it is not sufficient to observe a group of individuals who received the treatment; we also need to observe a group of individuals who did not receive the treatment.

Note that the "hat" on top of the name denotes that this is an estimate, that is, a calculation based on approximations. All estimates, including this one, contain some uncertainty. (We will see how to quantify this uncertainty in chapter 7.)

It is worth repeating that the difference-in-means is a valid estimator of the average causal effect of a treatment on an outcome only when the treatment and control groups are comparable with respect to all the variables that might affect the outcome other than the treatment variable itself. As stated earlier, this is best achieved in experiments such as Project STAR, in which the treatment is randomly assigned. The randomization of treatment assignment enables researchers to isolate the effect of the treatment from the effects of other factors.

> ESTIMATING AVERAGE CAUSAL EFFECTS USING RANDOMIZED EXPERIMENTS AND THE DIFFERENCE-IN-MEANS ESTIMATOR: By using random treatment assignment, we can assume that the treatment and control groups were comparable before the administration of the treatment. As a result, we can rely on the difference-in-means estimator to provide a valid estimate of the average treatment effect.

Unfortunately, we are not always able to conduct an experiment. Three types of obstacles might prevent us from running one:

- Ethical: It would not be ethical to randomize certain treatments, such as a potentially lethal drug.
- Logistical: Some treatments, such as height or race, cannot be easily manipulated.
- Financial: Experiments are often expensive. Project STAR cost many millions of dollars, for example.

Experimental data are data collected from a randomized experiment, whereas **observational data** are data collected about naturally occurring events. Studies that use observational data are called **observational studies**.

Given that we cannot always run experiments, we need to learn how to estimate causal effects in non-experimental settings, using what is called **observational data**. Unlike **experimental data**, which refers to data collected from a randomized experiment, observational data are collected about naturally occurring events. Treatment assignment is out of the control of the researchers and is often the result of individual choices. For example, we may want to estimate the average causal effect of small classes on student performance by collecting data from school districts where the size of the classes varies as a result of factors such as school budgets, student enrollment, or the physical limitations of the school buildings. In these types of studies, known as **observational studies**, we have to find a statistical way to make treatment and control groups comparable without relying on the randomization of treatment assignment. We will learn how to do this in chapter 5.

Now that we know that when analyzing the STAR dataset, we can use the difference-in-means estimator to estimate the average causal effect of small classes on student performance, it is time to perform the analysis.

2.5 DO SMALL CLASSES IMPROVE STUDENT PERFORMANCE?

To follow along with this chapter's analysis, you may create a new R script in RStudio and practice typing the code yourself. Alternatively, you may open "Experimental.R" in RStudio, which contains all the code for this chapter. We begin the analysis by running the following code from the previous chapter:

TIP: If you are starting a new R session, to operate with the data, you need to re-run some of the code we wrote in the previous chapter, specifically the lines of code that:

- set the working directory to the folder containing the dataset using the function setwd()
- read the dataset using read.csv() and store it as an object called *star* using the assignment operator <-.

We provide here the code to set the working directory if the DSS folder is saved directly on your Desktop. (Note that in the code for Windows computers, you must substitute your own username for *user*.) If the DSS folder is saved elsewhere, please see subsection 1.7.1 for instructions on how to set the working directory.

```
setwd("~/Desktop/DSS") # setwd() if Mac
setwd("C:/user/Desktop/DSS") # setwd() if Windows

star <- read.csv("STAR.csv") # reads and stores data

head(star) # shows first observations
##   classtype reading math graduated
## 1     small     578  610         1
## 2   regular     612  612         1
## 3   regular     583  606         1
## 4     small     661  648         1
## 5     small     614  636         1
## 6   regular     610  603         0
```

Here, we are interested in using this dataset to estimate the average causal effect of attending a small class on three different measures of student performance: *reading*, *math*, and *graduated*. For each outcome variable, we need to perform a separate analysis. Since Project STAR was a randomized experiment, we can use the difference-in-means estimator to estimate each of the three average treatment effects.

Before we can compute the difference-in-means estimators, we need to learn to use relational operators, which enable us to create and subset variables.

2.5.1 RELATIONAL OPERATORS IN R

There are many relational operators in R that can be used to set a logical test. In this book, we use only the operator ==, which evaluates whether two values are equal to each other. If they are, R returns the logical value TRUE. If they are not, R returns the logical value FALSE. (TRUE and FALSE are not character values. They are special values in R, with a specific meaning, and therefore are not written in quotes.) For example, if we run:

== is the relational operator that evaluates whether two values are equal to each other. The output is a logical value: TRUE or FALSE. Example: 3==3.

```
3==3
## [1] TRUE
```

R lets us know that indeed 3 equals 3. If we instead run:

```
3==4
## [1] FALSE
```

R returns a FALSE, indicating that 3 is not equal to 4.

RECALL: In the STAR dataset, the variable *classtype* identifies the class the student attended. In R, we use the $ character to access a variable inside a dataframe. To its left, we specify the name of the object where the dataframe is stored (without quotes). To its right, we specify the name of the variable (without quotes). Example: *data*$*variable*. We use quotes " around values that are text but not around values that are numbers. In the output, the numbers in brackets at the beginning of each line indicate the position of the observation immediately to the right.

We can apply relational operators to all the values in a variable at once. In this case, R considers the value of each observation one by one and returns a TRUE or a FALSE for each of them. For instance, if we wanted to determine which students in the STAR dataset attended a small class, we run:

```
star$classtype=="small"
## [1]  TRUE FALSE FALSE TRUE TRUE FALSE
## [7]  TRUE TRUE FALSE FALSE FALSE FALSE
##  ...
```

After running the code above, R returns as many logical values as observations in the variable *classtype*. (Here we show you only the first 12.) For students who attended a small class, R returns TRUE because the value of *classtype* equals "small". For students who did not, R returns FALSE. For example, as we saw in the output of head() above, the value of *classtype* for the first observation is "small", and therefore, here R returns TRUE as the first output.

Now we can ask R to perform a different action depending on the results from a logical test (the TRUE or FALSE returned from applying the == operator). For example, we can ask R to produce values for a new variable or to extract specific values from an existing variable based on the results of the logical test.

2.5.2 CREATING NEW VARIABLES

Using the function ifelse(), which stands for "if logical test is true, return this, else return that," we can create the contents of a new variable based on whether the values of an existing variable pass a logical test. For example, we can create the contents of a new binary variable based on the values of *classtype*. For the students whose value of *classtype* equals "small", we ask R to return a 1, and for all other students a 0.

ifelse() creates the contents of a new variable based on the values of an existing one. It requires three arguments in the following order, separated by commas: (1) the logical test, (2) return value if test is true, and (3) return value if test is false.

== is the relational operator we use to set the logical test that evaluates whether the observations of a variable are equal to a particular value. We write values in quotes " if text but not if numbers.

Example: ifelse(*data*$*var*=="yes", 1, 0) returns a 1 when *var* equals "yes" and a 0 otherwise, creating the contents of a binary variable using the existing character variable *var*.

The function ifelse() requires three arguments:

- The first is the logical test, which specifies the true/false question that serves as the criterion for creating the contents of the new variable. In the current application, for every student, we want to evaluate whether the value of *classtype* equals "small". As shown above, the code star$classtype=="small" accomplishes this.
- The second argument is the value we want the function to return when the logical test is true. In this case, we want the return value to be a 1 whenever *classtype* equals "small".

- The third argument is the value we want the function to return when the logical test is false. In this case, we want the return value to be a 0 whenever *classtype* does not equal "small".

Go ahead and run the following code:

```
ifelse (star$classtype=="small", 1, 0)
## [1] 1 0 0 1 1 0 1 1 0 0 0 0
## ...
```

TIP: Here, the first return value is a 1 and the second is a 0. Why? In the first observation of the STAR dataset, *classtype* equals "small", and so the logical test is TRUE, and therefore, the ifelse() function returns a 1. In the second observation, *classtype* equals "regular", and so the logical test is FALSE, and therefore, the ifelse() function returns a 0.

The function returns a 1 or a 0 for every student in the STAR dataset depending on the type of class they attended. (Here again, we show you only the first 12 values.)

To store these values as a new variable, we use the assignment operator <-. To its left, we need to specify the name of the new variable. Here, we chose to name the variable *small*. To store it as a variable inside the dataframe and not just as a new object by itself, we need to identify the name of the dataframe before the name of the variable with the $ character in between. (Note that the $ character allows us to create a new variable, and not just access an existing one as we saw in chapter 1.)

$ is the character used to identify a variable inside a dataframe, either to access it or to create it. To its left, we specify the name of the object where the dataframe is stored (without quotes). To its right, we specify the name of the variable (without quotes). Example: *data$variable*.

Putting it all together, to create the new variable *small* we run:

```
star$small <- ifelse (star$classtype == "small", 1, 0)
```

TIP: Recall that the name of an object or variable can be anything as long as it does not begin with a number or contain spaces or special symbols like $ or %. For practical reasons, the name of an object or variable should reflect the meaning of its contents, be short, and be written in all lowercase letters.

Whenever you create a new variable, it is good practice to check its contents. Doing so can save you a lot of trouble down the road. For example, here we can take a quick look at the first few observations of the dataframe using head() to ensure that the new binary variable was created correctly.

```
head(star) # shows first observations
##   classtype reading math graduated small
## 1     small     578  610         1     1
## 2   regular     612  612         1     0
## 3   regular     583  606         1     0
## 4     small     661  648         1     1
## 5     small     614  636         1     1
## 6   regular     610  603         0     0
```

Looking at the output, we can see that we have a new variable called *small*. Comparing the values of *small* to the values of *classtype*, we can confirm that whenever *classtype* equals "small", *small* equals 1 and that whenever *classtype* equals "regular", *small* equals 0. Indeed, in the first, fourth, and fifth observations, the value of *classtype* is "small" and the value of *small* is 1. In the second, third, and sixth observations, the value of *classtype* is "regular" and the value of *small* is 0.

2.5.3 SUBSETTING VARIABLES

`[]` is the operator used to extract a selection of observations from a variable. To its left, we specify the variable we want to subset. Inside the square brackets, we specify the criterion of selection. For example, we can specify a logical test using the relational operator `==`. Only the observations for which the logical test is true will be extracted. Example: `data$var1[data$var2==1]` extracts the observations of the variable *var1* for which the variable *var2* equals 1.

Using square brackets `[]`, we can extract the selection of observations for which a logical test is true. This is useful in a variety of situations. For example, to estimate the average causal effect of *small* on *reading*, we need to compute the following difference-in-means estimator:

$$\begin{pmatrix}\text{average reading}\\ \text{test scores among}\\ \text{students in}\\ \text{small classes}\end{pmatrix} - \begin{pmatrix}\text{average reading}\\ \text{test scores among}\\ \text{students in}\\ \text{regular-size classes}\end{pmatrix}$$

This formula requires calculating the averages of two subsets of observations of *reading* for which a certain criterion is met. To subset a variable, we use the `[]` operator. To its left, we specify the variable we want to subset, `star$reading` in this case. Inside the square brackets, we specify the criterion of selection. The examples below should clarify how this works.

RECALL: `mean()` calculates the mean of a variable. The only required argument is the code identifying the variable. Example: `mean(data$variable)`.

As stated in the previous chapter, we can use the function `mean()` to compute the mean of a variable in R. To calculate the average reading scores among all students in the STAR dataset, we run:

```
mean(star$reading) # calculates the mean of reading
## [1] 628.803
```

To calculate average reading scores among *only* the students who attended a small class, we need to include in the average *only* the observations of *reading* for which *small* equals 1. The following code accomplishes this:

```
mean(star$reading[star$small==1]) # for treatment group
## [1] 632.7026
```

Values of *small* and *reading* for the first six observations in the STAR dataset. Observations from students who attended a small class (*small*=1) are in black, and observations from students who attended a regular-size class (*small*=0) are in gray.

i	*small*	*reading*
1	1	578
2	0	612
3	0	583
4	1	661
5	1	614
6	0	610

Only the observations of *reading* for which the logical test specified inside the square brackets is true are selected for the computation of the mean. For example, among the first six observations in the dataset, only the values of *reading* that correspond to observations 1, 4, and 5 are included in this average. (See the table in the margin.) According to the output above, students who attended a small class earned about 633 points on the reading test, on average.

How about the students who attended a regular-size class? The code to compute this mean is identical to the one above, except that now the criterion of inclusion is that *small* must equal 0.

```
mean(star$reading[star$small==0]) # for control group
## [1] 625.492
```

Based on this output, students who attended a regular-size class earned about 625 points on the reading test, on average.

Now we can easily calculate the difference-in-means estimator as the difference between these two averages using the outputs above (633 − 625). Better yet, we can compute it all at once, by running the following piece of code:

```
## compute difference-in-means estimator for reading
mean(star$reading[star$small==1]) -
  mean(star$reading[star$small==0])
## [1] 7.210547
```

For the other two outcome variables, then, we can compute the corresponding difference-in-means estimators as follows:

```
## compute difference-in-means estimator for math
mean(star$math[star$small==1]) -
  mean(star$math[star$small==0])
## [1] 5.989905

## compute difference-in-means estimator for graduated
mean(star$graduated[star$small==1]) -
  mean(star$graduated[star$small==0])
## [1] 0.007031124
```

These two pieces of code are identical to the previous one, except that now we use *math* and *graduated*, respectively, instead of *reading* as the outcome variable of interest.

What can we conclude from these results? Assuming that the students who attended a small class were comparable before schooling to those who attended a regular-size class (a reasonable assumption given that the dataset comes from a randomized experiment), we estimate that attending a small class:

- increased student performance on the third-grade reading test by 7 points, on average
- increased student performance on the third-grade math test by 6 points, on average
- increased the proportion of students graduating from high school by about 1 percentage point, on average.

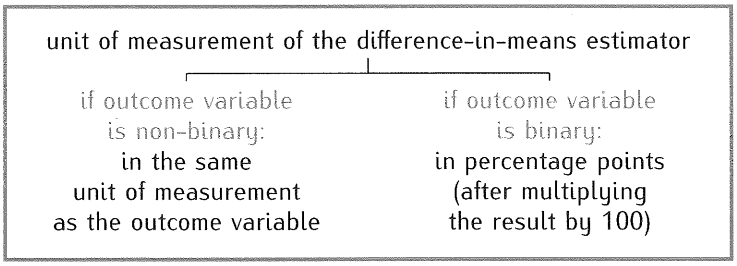
Notice that conclusion statements should mention the key elements of the analysis. (See TIP in the margin.) In addition, note that the unit of measurement of the difference-in-means estimator differs depending on the type of outcome variable. See the summary provided in outline 2.1. (Just as we did when discussing the interpretation of means in chapter 1, we exclude categorical variables from this discussion.)

TIP: By convention, when we include in the R script a comment at the beginning of a line, as opposed to after some code, we use two # characters instead of one.

TIP: Good conclusion statements are clear, are concise, and include the key elements of the analysis. For example, when estimating average causal effects with randomized experiments, be sure to convey:

- the assumption: the treatment and control groups are comparable based on pre-treatment characteristics; in this case, students who attended a small class were comparable before schooling to those who attended a regular-size class
- the justification for the assumption: dataset comes from a randomized experiment
- the treatment: attending a small class
- the outcome variable(s): third-grade reading test scores, third-grade math test scores, and proportion of students graduating from high school
- the direction, size, and unit of measurement of the causal effect(s): an increase of 7 points, an increase of 6 points, and an increase of a little less than 1 percentage point, respectively
- the fact that you are making a causal claim: use causal language (attending a small class *increased* student performance) rather than observational language (students attending a small class performed better than students attending a regular-size one)
- the fact that you are estimating *average* causal effects as opposed to individual causal effects.

OUTLINE 2.1. Unit of measurement of the difference-in-means estimator based on the type of outcome variable.



unit of measurement of the difference-in-means estimator

if outcome variable is non-binary: in the same unit of measurement as the outcome variable

if outcome variable is binary: in percentage points (after multiplying the result by 100)

If the outcome variable is non-binary, the unit of measurement of the difference-in-means estimator will be the same as the unit of measurement of the outcome variable. For example, if the outcome variable is measured in points, as is the case with both *reading* and *math*, then the average outcomes for the treatment and control groups will also be in points (the average of points is measured in points) and so will be the estimator (points−points=points).

TIP: What is a percentage point? It is the unit of measurement for the arithmetic difference between two percentages. For example, if a student's proportion of correct answers on a test improved from 50% to 60%, we would state that the score increased by 10 percentage points:

$$\begin{aligned}\triangle\text{score} &= \text{score}_{\text{final}} - \text{score}_{\text{initial}}\\ &= 60\% - 50\% = 10 \text{ p.p.}\end{aligned}$$

Why is this difference not referred to as 10%? Because percentage change is different from percentage-point change. If someone told us that the initial score was 50% and that it increased by 10%, the final score would be 55% (not 60%). Because an increase of 10% of 50% is an increase of 5 percentage points (0.10×50=5 p.p.), the final score would be:

$$\begin{aligned}\text{score}_{\text{final}} &= \text{score}_{\text{initial}} + \triangle\text{score}\\ &= 50\% + 5 \text{ p.p.} = 55\%\end{aligned}$$

If the outcome variable is binary, the unit of measurement of the difference-in-means estimator will be percentage points, sometimes abbreviated as p.p. (after multiplying the output by 100). Why?

- First, as explained in the previous chapter, the average of a binary variable should be interpreted as a percentage (after multiplying the output by 100), because it is equivalent to the proportion of the observations that have the characteristic identified by the variable. As a result, when the outcome variable is binary, as is the case with *graduated*, the average outcomes for the treatment and control groups will both be measured in percentages (after multiplying the output by 100).
- Second, the unit of measurement for the arithmetic difference between two percentages is percentage points (percentage−percentage=percentage points). (See TIP in the margin.) Therefore, if the outcome variable is binary, the difference-in-means estimator will be measured in percentage points (after multiplying the output by 100).

As an example, let's revisit the interpretation of the difference-in-means estimator for the binary variable *graduated*.

First, calculate the average of *graduated* for students attending a small class and for students attending a regular-size class, separately:

```
mean(star$graduated[star$small==1]) # for treatment group
## [1] 0.8735043

mean(star$graduated[star$small==0]) # for control group
## [1] 0.8664731
```

The top output above indicates that among students who attended a small class, the average high school graduation rate was 87.35% (0.8735×100=87.35%). The bottom output indicates that among students who attended a regular-size class, the average high school graduation rate was 86.65% (0.8665×100=86.65%).

Second, compute the difference-in-means estimator, which is the difference between the two averages above:

```
## difference-in-means for graduated
0.8735043 - 0.8664731
## [1] 0.0070312
```

As we already knew from our calculations above, the difference-in-means estimator for *graduated* equals 0.007. It should be interpreted as an increase in the probability of graduating from high school of 0.7 percentage points, on average (0.007×100=0.7 p.p. or 87.35%−86.65%=0.7 p.p.).

TIP: Because an average causal effect estimates the *average change* in Y caused by a change in X, it should be interpreted as an average increase in Y if positive, as an average decrease in Y if negative, and as no average change in Y if zero.

Now that we have clarified how to interpret the difference-in-means estimator, let's return to our estimates of the average treatment effects above. There are two caveats to these estimates:

- First, they indicate how much the average outcome across multiple individuals changes as a result of the treatment. They do not indicate how the treatment would affect any one individual's outcome. As we saw in the idealized scenario earlier in the chapter, individual-level treatment effects might differ significantly from average treatment effects. While we estimate that student performance on the reading test improved, on average, as a result of attending a small class, a particular student's performance might have suffered from it.
- Second, the validity of these estimates rests on the plausibility of the assumption that the treatment and control groups are comparable with respect to all the variables that might affect the outcome other than the treatment variable itself. In this case, we can confidently make this assumption because we are analyzing data from a randomized experiment.

There are still a few questions that we need to answer to complete this analysis. Two in particular are worth noting here:

- Can we generalize these results to a population of students other than those who participated in Project STAR?
- Do the estimated causal effects represent real systematic effects rather than noise in the data?

We learn how to answer the first type of question in chapter 5 and explore the second in chapter 7, once we have become acquainted with the relevant concepts.

2.6 SUMMARY

In this chapter, we learned about causal effects and some of the difficulties we face when attempting to estimate them.

If we could observe the outcomes of the same individual under both treatment and control conditions at the same time, we could compute the causal effect of the treatment on a particular individual's outcome as the difference between these two potential outcomes.

Unfortunately, observing both potential outcomes is not possible. In reality, we observe only the outcome under the condition each individual received (the factual outcome) and can never observe what would have happened had the individual received the opposite condition (the counterfactual outcome).

To estimate a causal effect, we have to rely on assumptions to approximate the counterfactual outcome. This leads us to estimate average treatment effects across multiple individuals rather than the treatment effect for each individual.

When the treatment and control groups are comparable, we can use the average observed outcome (the factual outcome) of one group as a good approximation for the average unobserved outcome (the counterfactual outcome) of the other. Under these circumstances, the difference-in-means estimator produces a valid estimate of the average treatment effect.

The best way of ensuring that treatment and control groups are comparable is to run a randomized experiment. By assigning individuals to the treatment or control group based on a random process such as a coin flip, we ensure that the two groups have identical pre-treatment characteristics, on average. Later in the book, we will learn how to estimate average causal effects when we cannot run a randomized experiment and, instead, must analyze observational data.