CHAPTER 2

Foundational Concepts for Quantitative Research

Learning Objectives

After reading this chapter, you will be able to do the following:

1. Define basic terms for quantitative research.
2. Describe the research circle.
3. Identify the four major goals of social research.
4. Write a checklist of the W's.
5. Understand the reasons for both reporting and interpreting numbers.
6. State the importance of specifying the direction and magnitude of a pattern.

In this chapter, I lay the groundwork for the rest of the book by defining and illustrating some basic research methods terminology. I then describe the research circle and the strategies it encompasses and identify the four major goals of research. Finally, I introduce some basic principles for working with numeric information.

Terminology for Quantitative Research

To begin learning how to make sense of numbers, you need to become familiar with the terminology that researchers use when discussing quantitative data and research methods. The term **data** refers to information that has been collected on characteristics such as age, educational attainment, and place of residence of individual people or information on type of cuisine, price range, and customer satisfaction ratings for individual restaurants.

Utts (1999) makes a useful distinction between **data** and **numbers**, stating that "data are numbers to which meaning has been attached" (p. 15). **Data** is the plural form of the singular **datum**, which is why we write "data are" rather than "data is."

Example: "2" is a number, but 2 people in a family or $2 for a cup of coffee are data.

*The numeral "2" in the first phrase is a **number** (not data) because it isn't associated with a topic or units. When used to describe family size, it has a **topic** (family size) and **units** (number of people), so that same **numeral** is now **data**. "2" is also **data** when it pertains to price of coffee (**topic**) and $ per cup (**units**).*

Raw data refers to category labels (such as type of religion or name of country) or numeric information that has been collected but not yet processed or manipulated in any way (Utts & Heckard, 2014, p. 15). To turn raw data into useful evidence, those data have to be organized, summarized, and compared with other values to help interpret what they mean for the question at hand. We'll see some introductory examples in the section on Reporting and Interpreting Numbers.

An **observational unit** is an individual **entity** (such as a person, institution, or geographic area) about whom data are collected. However, for studies that collect information on the same cases across time (longitudinal studies; Chapter 12), each data collection time point is called an **observation**. In those situations, each observational unit is referred to as a **respondent** (for studies of people), **case**, or **study subject**. (Note that this usage of "study subject" refers to study participants, not to the topic of the study.)

Example: Consider two different studies: A survey of college students and a study of family leave policies at different corporations.

*In the survey of college students, the type of **entity** is a student, and each student is an **observational unit** (or **study subject**). In the family leave study, the type of **entity** is a corporation, and each corporation is a **case**.*

Example: For a longitudinal study of how tree growth is affected by climate change, scientists measure the diameter of each tree in their study plot every year for 20 years.

*Each individual tree is a **case** and each annual measurement of tree diameter is an **observation**.*

Data are information that has been collected about characteristics of **entities**. An **observational unit** (sometimes called a **case, respondent**, or **study subject**) is an individual **entity** about which data are collected, such as people, neighborhoods, or schools. These are also known as **units of analysis** or **levels of aggregation**. In studies that follow cases across time, there are multiple **observations** for each case.

A **variable** is a characteristic that can vary from one case to another. Each case has a **value** for each variable, capturing the nature or extent of that characteristic for that case. In contrast to a variable, a **constant** is a characteristic that has the same value for all cases being studied, often due to restrictions on which cases are included in the data set.

> <u>Example:</u> In 2020, Canada was a democracy, whereas North Korea was a communist country.
>
> > *Type of political system is a **variable**, but date is a **constant**. The type of **entity** is a nation.*
>
> <u>Example:</u> Heights of each child in a kindergarten class are measured. Maria is 115 centimeters (cm.) tall.
>
> > *For each of her classmates, the **variable** height will take on the **value** that captures their height—e.g., 112, or 115, or 119 cm. Every student in Maria's class is a human being, so in that context, "species" is a **constant**.*

However, which characteristics are constants and which are variables depends on the topic and context.

> <u>Example:</u> A study of the Metropolitan zoo collects data on all the different animals, including information on their ages.
>
> > *In this study, "species" is a **variable** because some of the zoo residents are lions, some are eagles, some are rattlesnakes, and so on. Age is also a **variable**, with some young and some older animals. However, all of the zoo residents are animals, so in that study, type of biological **entity** is a **constant**.*
>
> <u>Example</u>: In Maria's kindergarten class, every student is 5 years old at the beginning of the school year.
>
> > *For that topic, setting, and time point, age is a **constant**. However, if the study also included other grades within her elementary school or followed the kids in her class for several years, age would be a **variable**.*

A **data set** is an organized collection of information on a consistent set of variables for each of the cases in a study. **Missing values** occur when a respondent did not provide information on a particular variable. In longitudinal studies, there may be some variation in the amount of information collected for different participants, based on how often each of them participated.

> <u>Example</u>: A social survey asks each of 1,000 respondents the same set of questions about their demographic traits, volunteer activities, and attitudes about current policy issues. Some of the study participants did not answer the question about a proposed health care policy.
>
> > *A **data set** for that survey would contain answers to each of those questions for each **respondent**. In other words, that survey does <u>not</u> have*

one set of questions for some respondents and a completely different set of questions for others. Participants who didn't answer the policy question would have a ***missing value*** *for that* ***variable****.*

Example: An online shopping website collects data on each customer's shipping address, searches, and purchases.

Location, search topics, items purchased, and dates of purchase are some of the ***variables*** *on which data were collected from every customer (****case****) who shopped on that site. The* ***data set*** *will include more information on each of those topics for people who visited the site several times than for those who visited the site only once.*

A **variable** is a characteristic or attribute that can vary (differ) from one case to another. Each case takes on a **value** for a variable, measuring the type or degree of that characteristic for that case. A **constant** is a measure that has the same value for every case in the context under study. A **data set** or **database** is a compilation of information on the same set of variables for each of the cases under study. Cases that do not provide information on a particular variable have a **missing value** for that variable.

Figure 2.1 is a schematic diagram of a hypothetical school health data set, with one row for each student in the school and one column for each attribute about which data were collected. For each of the 432 cases (students in the school), the data set includes information on each of the variables (attributes). The ". . ." in the row between ID 105 and ID 432 shows that there were additional students in the data set, even though they are not included in the diagram.

Figure 2.1 Hypothetical Data Set of School Health Records

ID	School code	Date of medical record	Age (years)	Gender	Height (cm.)	Weight (kg.)	Date of most recent vaccination	Transfer status	Code of prev. school
101	1227	9/15/18	5	M	117	20	9/15/18	No	NA
102	1227	9/2/18	8	M	129	26	7/10/17	No	NA
103	1227	8/27/18	6	F	119	19	3/22/17	Yes	3418
104	1227	8/28/18	10	M	140	33	11/2/17	No	NA
105	1227	9/7/18	8	F	123	25	4/30/18	No	NA
. . . .		. . .							
432	1227	9/15/18	7	F	118	23	12/5/16	Yes	5009

Example: Student 101 is a 5-year-old male who was 117 centimeters tall and weighed 20 kilograms when he visited his doctor September 15, 2018. His most recent vaccination took place on that same date.

*That sentence reports the **case's** values for the **variables** age, gender, height, weight, and dates of medical record and vaccination. Those same **variables** take on different **values** for other **cases** (students) in the **data set.***

Example: The ages of the students in School 1227 range from 5 to 10 years; two of the cases shown are both 8 years old.

*In large data sets, many cases can share the same **value** of a **variable**—in this example, age. School code is a **constant** for every student in the data set because they all attend the same school. If the study included more than one school, then school code would be a **variable** because different students would come from different schools.*

If certain questions pertain to only some of the cases in a data set, those questions will be not be asked about those for whom they do not apply. We'll learn more about missing values and "not applicable" responses in Chapter 4.

Example: Students 103 and 432 both transferred in from other schools midyear, whereas the other students attended School 1227 from the beginning of the academic year.

*As shown in the two right-hand columns of Figure 2.1, for the students who transferred to School 1227, the **value** of the transfer status **variable** is "yes," and the "previous school" variable has a code for another specific school. For the students who attended School 1227 from the beginning of the year, the value of the transfer status variable is "no," and the "previous school" variable is coded "NA" for **not applicable**. By indicating the cases to which a particular question does not apply, the **data set** can have a consistent list of variables for every case, with their respective values filled in.*

The Research Circle

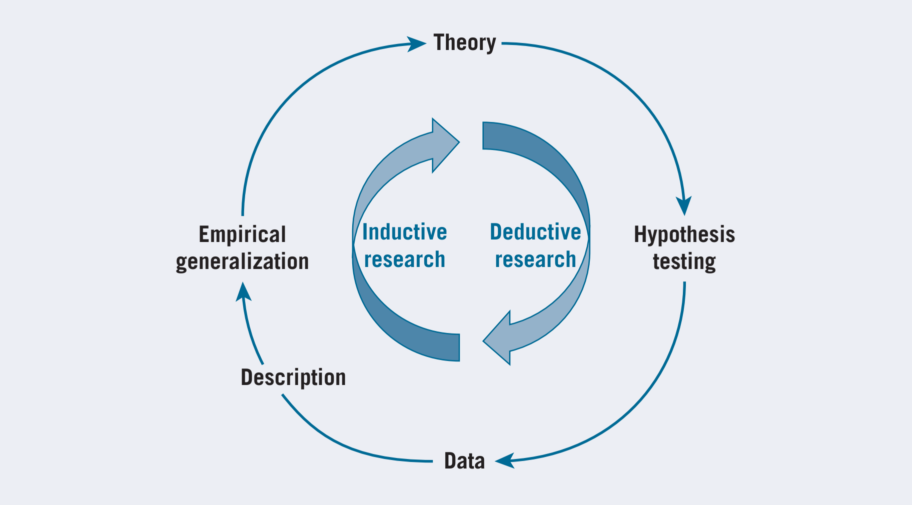
A wide variety of types of methods can be used to conduct research: **Quantitative methods** use numbers as evidence, collecting information from sources such as surveys, censuses, surveillance data, and "big data" (Chapter 11). By definition, quantitative methods can only be used to study phenomena that can be quantified in some way, whether by counting things or measuring their dimensions or characteristics. **Qualitative methods** are used to study phenomena that cannot easily be quantified, working with data collected using focus groups, participation/observation, or in-depth interviews or sources such as audio, video, or text documents. **Mixed-methods** research uses a combination of quantitative and qualitative approaches within a research study or set of related studies to take advantage of their particular strengths.

With its objective of teaching how to make sense of numbers, this book focuses on quantitative research methods. To learn more about qualitative research methods, see Chambliss and Schutt (2018) or Mack et al. (2005).

> **Quantitative research methods** are used to study phenomena that can be counted or measured numerically. **Qualitative research methods** are used to study phenomena that cannot easily be quantified. **Mixed-methods research** uses a combination of quantitative and qualitative approaches.

There are two major types of research strategies: **Inductive** and **deductive**. **Descriptive research** is an intermediary step that involves making empirical generalizations based on the data, where **empirical** refers to observed data rather than to ideas based solely on theory or logic. The overall research process combines descriptive, inductive, and deductive research to create the **research circle** shown in Figure 2.2. As the circle implies, the research process does not have any built-in starting or ending point: A researcher can start at any point on the circle moving from theory to data (deductive—the arrow that starts at the top of Figure 2.2) or from data to theory (inductive—starting at the bottom of the circle), depending on the objective of their study.

Figure 2.2 The Research Circle



Source: Chambliss & Schutt (2016). Reprinted with permission.

Inductive Research

Inductive research starts with data and uses it to generate a theory about observed patterns. Put differently, a theory is "induced" (proposed) to account for patterns in the data. In everyday conversation, the word **theory** is used to simply refer to an idea, whether or not it has a factual basis. In research terminology, however, it has a much more formal and specific meaning: a hypothetical explanation of some aspect of the world that has been developed from observing data and describing patterns.

> <u>Example:</u> A study of drunk driving started with descriptive data showing that the rate of alcohol-related traffic accidents in the boroughs of New York City decreased by 24% to 35% after introduction of ride-hailing apps such as Uber, but only in the four boroughs where such apps were rapidly adopted (Peck, 2017). The author then developed a theory based on that pattern, surmising that availability of ride-sharing made it easier for people to avoid driving drunk.
>
> > *This research was **inductive** because it started with data on alcohol-related traffic accidents and **described** that the accident rate dropped after introduction of ride-hailing apps. That generalization **induced** the **theory** about a possible mechanism linking increased availability of ride-sharing services to observed (**empirical**) reductions in traffic accident rates.*

Inductive research is often conducted using qualitative or descriptive quantitative methods.

Deductive Research

Deductive research develops a hypothesis from theory and tests that hypothesis using data. In other words, a specific expectation about a relationship between two variables is **deduced** based on (reasoned from) a theory. A **hypothesis** is an educated guess or supposition about a relationship between two or more variables, specifying how the researcher anticipates the value of one variable will differ or change as some other variable differs or changes. A hypothesis should be developed based on a combination of theory and empirical evidence from descriptive studies or from previous research on the topic under study. Deductive research is often conducted using quantitative research methods.

> <u>Example</u>: A study of alcohol-related car crashes might start with a theory about the influence of ride-hailing apps and from it develop a hypothesis that greater availability of ride sharing is responsible for the observed decline in alcohol-related crashes.
>
> > *This research is **deductive** because it starts with theory and develops a **hypothesis**.*

The **research circle** shows how descriptive, inductive, and deductive research relate to data, theory, and hypotheses. **Empirical** means verifiable by observation or experience rather than based on theory or logic alone. **Inductive research** starts with data and develops a theory about why the observed patterns occur. A scientific **theory** is a hypothetical explanation of a relationship between concepts. **Deductive research** starts with a theory and develops a hypothesis, which is then tested using data. A **hypothesis** is a tentative statement predicting how variables are associated with one another.

In a hypothesis about a cause-and-effect relationship, the variable we think is the cause is referred to as the **independent variable**. We hypothesize that it predicts or explains variation in the dependent variable. The variable thought to be the effect (or **outcome**) is referred to as the **dependent variable**—so called because we hypothesize that it depends on, or changes in response to, variation in the independent variable. You will learn more about cause-and-effect relationships in Chapter 12 and how to test hypotheses in Chapter 13.

An **independent variable** is the characteristic that is thought to be the "cause" in a cause-and effect relationship. Synonyms for independent variable include **risk factor**, **predictor**, or **explanatory variable**. A **dependent variable** is the characteristic that is thought to be the "effect" in a cause-and effect relationship. Other commonly used terms for dependent variable are **outcome** or **response variable**.

Example: Keith et al. (2017) hypothesized that people with darker skin color would experience more frequent or severe racial discrimination.

Their hypothesis implies that extent of racial discrimination depends on skin color, so skin color is the ***independent variable*** *and racial discrimination is the* ***dependent variable***.

A hypothesis can be worded in an "if–then" format, implying that if the independent variable changes (or varies) in a certain way, then the dependent variable is expected to change (or vary) in a specified direction.

Example: If ride-sharing becomes more widely available, then alcohol-related traffic accidents will decrease.

Availability of ride-sharing services is the ***independent variable****, while traffic accident rate is the* ***dependent variable****. In other words, the* *hypothesis conjectures that traffic accident rates (the "effect") are influenced by availability of ride-share services (the "cause"). The Peck (2017) study did not have data on individuals' actual usage of ride-sharing, so a* ***deductive*** *study would need to collect such data to test this* ***hypothesis****.*

The decision of whether to use a descriptive, inductive, or deductive strategy for a particular study should be based on what is already understood about the topic and what additional questions have been raised by prior research or new circumstances. The overall body of research on a particular topic typically involves all three types of strategies to provide a more complete understanding of that topic. Some researchers specialize in either inductive or deductive research, complementing studies of the same topic that other researchers have conducted using a different research strategy. Other researchers conduct both inductive and deductive research over the course of their career, designing each study to fill in gaps in what they had learned from their own prior work and that of others studying that topic.

Goals of Quantitative Research

A research project can have any of several **goals**: descriptive, exploratory, explanatory, and evaluation. Any of these goals can be pursued using either quantitative or qualitative approaches, however, given the focus of this book, the examples below describe hypothetical quantitative studies.

Descriptive Research

Descriptive research is exactly what it sounds like: It seeks simply to describe a pattern or relationship without aiming to explain the reasons for or implications of that pattern; those are covered by one or more of the other research goals. Descriptive research starts by defining what is meant by each concept (aspect of the question) and determining how that attribute will be measured.

Example: A consulting company wants to ensure that their clients are satisfied with the services provided through their online portal. First, they decide to define "satisfaction" as encompassing ease of finding information, coverage of topics clients want to learn about, and accessibility and quality of their web customer support. They develop a survey to measure those dimensions and then analyze the data to identify which aspects of their portal need the most improvement.

This ***descriptive study*** *uses data collected via the survey to describe what percentage of clients have low satisfaction based on the way they defined and measured that concept.*

Descriptive research is often conducted when getting to know a new topic or to see how some phenomenon occurs in different locations, times, or groups. Other times, descriptive research is conducted to find out what can be learned about a topic using a different definition or measure of a concept. The observed patterns may then be used to generate theories about the relationship between the variables or to develop hypotheses to be studied with one of the other types of research. You will learn more about how to define concepts in Chapter 3 and how to specify approaches to measuring those concepts in Chapter 4.

Exploratory Research

Exploratory research seeks to understand how a phenomenon works, such as what problems people face in a certain social setting, how they cope, and how they think about their experiences. Findings of exploratory research can be used to identify concepts about which to collect data for future descriptive studies or to develop theories about relationships between the attributes of that setting.

> Example: A study of immigration collects data on where respondents came from, their reasons for leaving their home country, barriers they encountered along the way, and strategies they adopted to overcome those barriers.
>
> *This hypothetical* ***exploratory study*** *was designed to provide insight into what motivates people to emigrate, and to describe the problems and solutions they developed during their immigration experiences.*

Explanatory Research

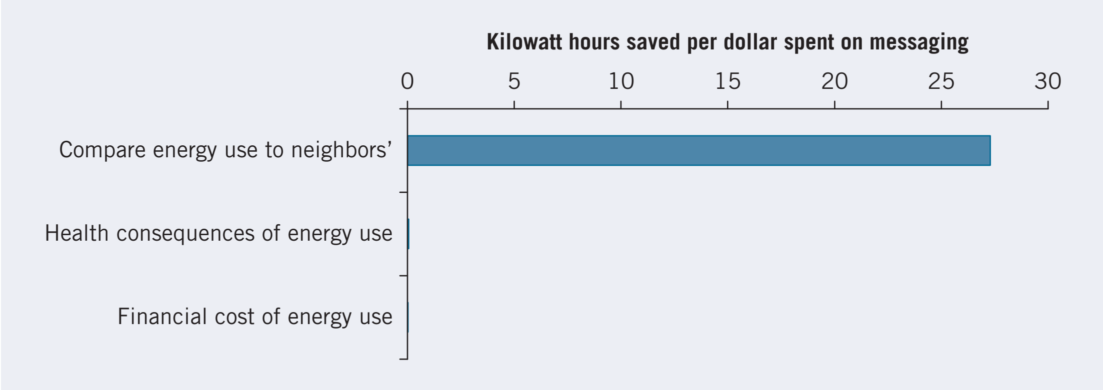
Explanatory research seeks to identify causes and effects of social, biological, or other phenomena and to predict how an outcome differs or changes in response to variation or change in some other characteristic. Explanatory research often involves formulating and testing a hypothesis about the relationship of interest. Results of explanatory research might then be used to design an intervention that could be tested with an evaluative study or inform a study to explore reasons for the observed patterns.

> Example: To test the theory that social comparison can be used as a "nudge" to reduce household energy consumption, researchers sent one of three messages to each of several thousand households as part of their monthly electric bill—either (1) comparing that household's energy consumption with their neighbors', (2) including a link to a website about the health consequences of pollution associated with energy consumption; or (3) pointing out the extra financial

cost to the household of higher electricity usage (Allcott & Rogers, 2014). They found that the households that received the "keeping up with the Joneses" message substantially reduced their electricity consumption, whereas those sent either of the other two approaches hardly changed their consumption at all (Figure 2.3; Benartzi et al., 2017).

*This **explanatory** study aimed to understand whether a social-comparative message (value of the **independent variable**) predicted greater change in household energy reduction (the **dependent variable**) than messages that did not involve social comparison.*

Figure 2.3 Comparison of Electricity Savings From Three Different Messages



Source: Benartzi et al. (2017) with data from Allcott & Rogers (2014). Reprinted with permission.

Evaluation Research

Evaluation research seeks to determine the effects of programs, policies, or other efforts to impact social, health or other outcomes, whether by government agencies, private nonprofits, or for-profit businesses.

<u>Example</u>: Concerned with rising obesity rates, a city imposes a tax on soda and other sweetened beverages. They compare obesity rates among their residents for the 3 years before and the 3 years after the tax was introduced.

*This study sought to **evaluate** whether <u>changing</u> the **independent variable** (presence of the soda tax) was associated with a change in the **dependent variable** (obesity rate).*

Descriptive research aims to define the concepts to be studied, and to characterize patterns in those phenomena. **Exploratory research** seeks to learn how a phenomenon works, such as how people get along in a particular social setting, what issues concern them, and what meanings they attach to their behaviors. **Explanatory research** strives to identify causes and effects of social or other phenomena and to predict how an outcome (dependent variable) varies or changes according to variation in some other characteristic (an independent variable). **Evaluation research** aims to assess the impact of programs, policies, or other interventions.

The topic of a study doesn't necessarily determine which type of research is appropriate. In fact, often, all four types of research will be conducted on a particular topic depending on what is already known about that topic and what remains to be learned about it. Suppose we want to learn how family socioeconomic status affects child well-being. Depending on the specific aspect of that relationship we are interested in learning about, any of the four research goals could be pursued.

Example: A study compares asthma rates for poor, near-poor, and nonpoor children, and reports what percentage of children are in each income group.

*Results of this **descriptive study** could be used to identify groups at high risk of asthma or to generate theories about the relationship between poverty and child health that could be used to develop hypotheses for an explanatory study.*

Example: A survey asks parents in low-income families what worries them about their children's health and safety and how they cope with low income while trying to provide for their children.

*This **exploratory study** is aimed at identifying possible reasons for patterns observed in prior descriptive studies of the relationship between poverty and child well-being. By learning the different ways parents in low-income families provide for their children, specific strategies could be identified to investigate in future studies using descriptive, explanatory, or evaluative approaches.*

Example: A study uses data on family income, what women ate while pregnant, and how much their babies weighed at birth to test whether inadequate intake of certain nutrients is the reason that low birth weight is more common among children from poor than nonpoor families.

*This **explanatory study** is aimed at figuring out what mechanisms explain the findings of a descriptive study showing that the percentage of babies that were low birth weight declined with increasing family income.*

Results of such a study could be used to design interventions to reduce low birth weight among children from low-income families.

Example: A study evaluates whether a pilot program to improve nutritional intake among children born into low-income families improves their health outcomes.

*This **evaluation** project tests whether an intervention to change the hypothesized cause (nutritional intake) results in better health outcomes for children in low-income families.*

Note that each of the four of the hypothetical studies is intended to learn more about the same topic—poverty and child well-being—but each type of research has a different goal in terms of what aspect of that topic is to be studied and why.

Several basic principles are central to making sense of numbers, affecting virtually every other quantitative reasoning skill you will learn in this book. They include using the W's as a checklist, reporting and interpreting numbers, and specifying the direction and magnitude of a pattern.

The W's

Information on **the W's** (who, what, when, where, and how) of a study is crucial for making sense of numbers. The W's are so essential that all of Part II of this book is devoted to explaining how those attributes affect many aspects of quantitative studies. Chapter 3 discusses the importance of the scope and definition of the topic: **what** is under study. Chapter 4 discusses the implications of **how** that concept was measured, and Chapter 5 covers the relevance of **when, where**, and to **whom** (or **which** cases) the data pertain. Later chapters cover other aspects of **how** the data were collected, analyzed, interpreted, and communicated as well as **how many** cases were studied.

> **The W's** consist of **what, when**, **where**, and to **whom/which** cases the data pertain; **how** data were defined, measured, collected, and analyzed, and **how many** (cases) are granted honorary W status because they, too, are vital to making sense of numbers.

For now, the key point is to pay attention to those W's for any numbers you work with, making note of them as you read about data that others have collected or analyzed, and writing about them as you present information about your own data collection or analysis.

The W's affect many aspects of how data are collected, analyzed, and interpreted and should be reported along with the associated numeric information.

Report and Interpret Numbers

In most quantitative research, numbers should be both reported and interpreted to help the audience understand the meaning of those numbers. **Reporting** numbers means including them in a sentence, table, or chart so readers can evaluate or analyze those values further. **Interpreting** numbers involves explaining what they mean in terms of the questions the author seeks to address, such as by comparing them to other numbers or explaining whether they are consistent with a hypothesis about the pattern under study. I introduce and define those concepts briefly here; we will return to many facets of reporting and interpreting numbers throughout the rest of this book.

Reporting the numbers is an important first step in communicating numeric information. By presenting (**reporting**) numbers in the text, table, or chart, researchers provide their audience with the raw data with which to perform additional calculations or comparisons to answer other questions of interest to them.

Example: A government report presents statistics on the inflation rate in their country for each of several years.

Readers could then use that numeric information to compare inflation rates with other countries or for other dates.

Example: A university website reports how many degrees it granted to men and women last year.

Readers who want to assess the size of the gender difference would be able to calculate various comparisons themselves, working from the numbers in the report.

Unless the sole purpose of a document is to make data available to potential users for their own analysis, however, authors should also **interpret** the numbers to help their readers make sense of the information. A number that has not been introduced or explained leaves it entirely to the audience to figure out what it means for the question under study. Those who are not familiar with the topic or setting are unlikely to know which comparisons to make or to have the information for those comparisons at hand. To help the audience understand what the numbers mean for that question and setting, authors should interpret (explain) the comparisons.

Example: A newspaper article reports last year's crime rate for a large city but does not describe whether that city's crime rate is trending up, down or is stable, or whether that city's crime rate is higher, lower, or similar to that in other cities.

The writer did not ***interpret*** *the crime rate, thus if readers want answers to those important questions, they would have to go to other sources to look up crime rates for other cities and years, and "do the math" themselves to compare them against the* ***reported*** *crime rate.*

Example: A middle-school teacher is planning an end-of-year trip to an amusement park that has just opened a new ride that all her students are talking about. If she doesn't compare her students' heights against the minimum height needed to go on that ride, she might have a lot of disappointed students on her hands!

*Here, knowing just the students' heights isn't enough. The teacher also needs to assess (****interpret****) whether most of her students are tall enough to participate.*

> **Reporting** numbers means presenting their values and the associated W's in the text, table, or chart. **Interpreting** means explaining how those numbers answer a question of interest.

Although it is important to interpret quantitative information, it is also essential to report the numbers. If a researcher only describes a ratio or percentage difference, for example, they will have painted an incomplete picture.

Example: A website advertises a sale offering 25% off all its products but doesn't list the prices themselves. A 25% reduction can be taken off prices at any level: ¥1.00, ¥400, or ¥2,000,000, to name just a few possibilities.

The percentage discount ***interprets*** *price differences without* ***reporting*** *the associated prices. Lacking data on either the original or discounted price on a particular product, customers won't be able to tell whether they can afford it even with the discount or how that store's discounted price compared to what the same product would cost at a different store.*

Example: Suppose a newspaper article reported that use of the Google Translate app doubled during the 2018 World Cup in Russia but did not report the amount translated either before or during that event. Doubling is consistent with many possible pairs of numbers: 100 translations before the World Cup versus 200 at the event, or 30,000 versus 60,000, or 2 million versus 4 million, for example.

*"Doubling" **interprets** the values by quantifying change over time. Unless the number of translations themselves are also **reported**, the popularity of the Google Translate app couldn't be compared with other ways of translating such as hardcopy phrase books or live translators. Furthermore, without info on the number of translations, it is impossible to tell whether that app was rarely used during the 2018 World Cup or was an essential tool for non-Russian speakers. The first pair of numbers would show that the app remained fairly unpopular even after the increase, the last pair that an already popular app went viral. In fact, the well-written article <u>did</u> **report** that the Google Translate app had "some 500 million users and translates some 143 <u>billion</u> words a day, into and out of dozens of languages," pointing out that even with that high baseline usage, the increase (doubling) of the Google Translate app was surprisingly large (Smith, 2018).*

> When numbers are intended as evidence to answer a question, they should be both reported and interpreted. When numbers are intended to mainly for others to use in their own calculations, the numbers may be reported but not interpreted.

Specify Direction and Magnitude

One of the most important aspects of interpreting a numeric comparison is describing both the direction and magnitude of that pattern.

Direction of Association

When comparing two or more values, **direction** identifies which of those values is higher. For change across time in the value of a variable, direction means conveying whether the trend is upward, level, or downward.

<u>Example</u>: "Today it is hotter in Vancouver than in Toronto,"

*"Hotter" captures the **direction** of the temperature difference between the two named locations.*

<u>Example</u>: "The average global temperature has risen steadily over the past few decades."

"Risen" expresses an <u>upward</u> temperature trend over time.

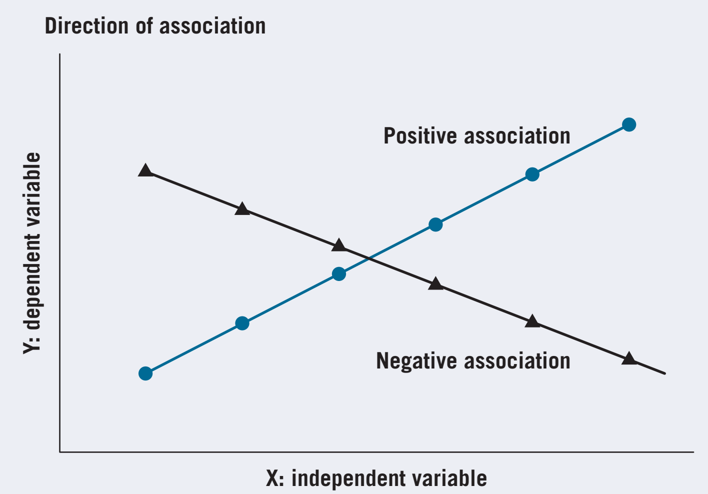
A relationship between two variables is called a **bivariate association**. There are two possible **directions of association** between quantitative variables. In a **positive association**, as the value of one variable increases, the value of the other variable also increases. A different way of explaining a positive association is that as the value of the independent variable decreases, the value of the dependent variable also decreases. Put differently, in a positive association, values of the two variables move in the same direction.

In a **negative association**, as the value of one variable increases, the value of the other variable decreases. In other words, in a negative association, the values of the two variables move in opposite directions. In math terms, direction specifies the **sign** (positive or negative) of the comparison.

Example: Children get taller with age.

*Age and height are **positively associated** because they move in the same **direction**, as in the blue line in Figure 2.4 that slopes upward from left to right.*

Figure 2.4 Positive and Negative Associations



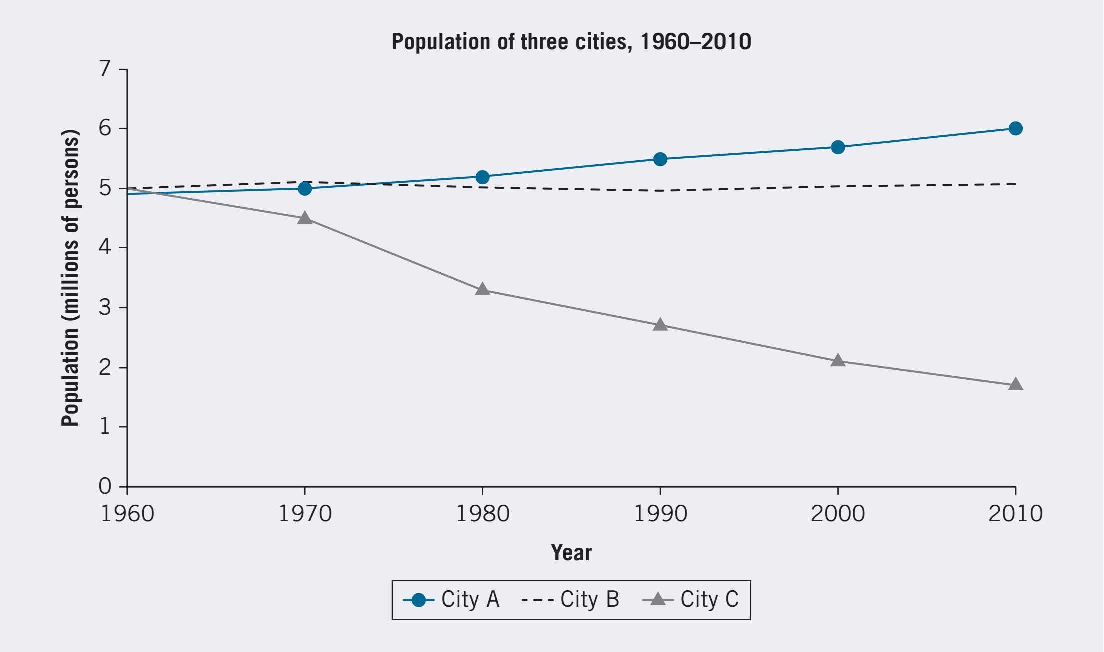
Example: As the price of an item increases, consumers buy less of that item.

*Price and quantity bought are **inversely** (**negatively**) associated because as one goes up, the other goes down, as in the downward-sloping black line in Figure 2.4.*

Example: The population of City A grew between 1960 and 2010, whereas that in City C decreased over that period (Figure 2.5).

*The **association** between time and population of City A is **positive**, while that in City C is **negative**.*

Figure 2.5 Illustration of Direction and Magnitude of a Trend



It is also possible to describe the shape of a **univariate** pattern, which looks at how values of one variable are distributed. You will learn more about univariate distributions in Chapter 10.

Example: In Figure 2.1, three of the listed students are male and three are female, so the two groups are equally common. One-third of the listed students transferred into School 1227 midyear, so they were less common than those who did not transfer in.

*Gender **composition** is **univariate** because it looks at how values of that one variable are distributed across the gender categories, without considering how it is related to any of the other variables in the data set. Likewise, the **transfer status** distribution conveys information about that one variable alone, so it is also **univariate**. "Less common" conveys **direction** by identifying which transfer status category is smaller than the other.*

> A **univariate distribution** (also known as **composition**) portrays how values of one variable are spread across possible values; "uni" means one, "variate" refers to a variable. A **bivariate association** examines how two variables are related to one

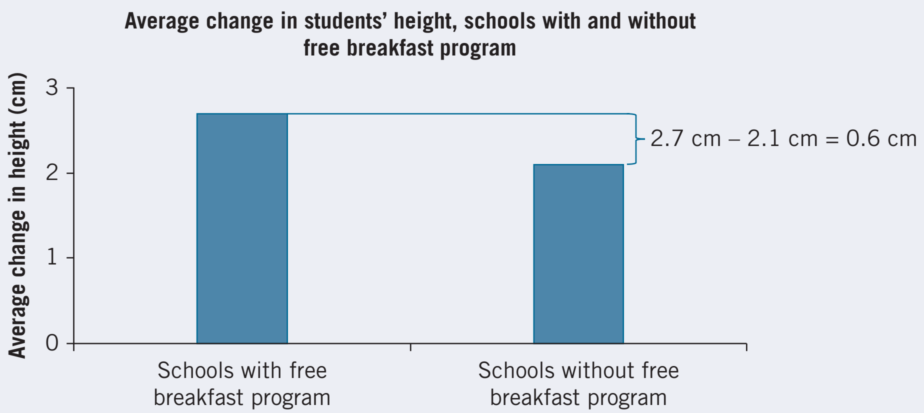
another; "bi" means two. The **direction** of an association conveys which of the values being compared is larger. In a **positive association** (also known as a **direct association)**, as the value of one variable increases, the value of the other variable also increases. In a **negative** (or **inverse**) **association**, as the value of one variable increases, the value of the other variable decreases.

Magnitude of an Association

Whereas direction identifies which value is bigger, **magnitude** captures how much bigger—in other words, the **size** of the difference between those values. We will learn various ways of calculating magnitude in Chapter 9 and how to use wording to express magnitude in Chapter 14.

> Example: Suppose a school district conducted an evaluation of whether a free breakfast program improves the nutritional status of students from low-income families. Half the elementary schools in the district were randomly assigned to provide free breakfasts to their low-income students and the other half to not provide free breakfast. Students were measured at the beginning and end of the school year and the average increases in height compared for the two groups of schools (Figure 2.6). At the end of the year, the researchers concluded that "children in the free breakfast program grew 0.6 centimeters more during the school

Figure 2.6 Illustration of Direction and Magnitude of an Association



Change in height from beginning to end of the school year.

year than those who did not receive free breakfasts (2.7 cm. and 2.1 cm. growth, respectively; Figure 2.7)."

*Students in both groups grew taller (same **direction** of change over the course of the year), but the free-breakfast group grew faster (larger **magnitude** of change). The word "more" conveys the **direction** of the association between program participation and growth in height), whereas "0.6 centimeters" (shown with the blue bracket) conveys the **magnitude** of that difference. This description both **reports** the information (the average height change for each group) and **interprets** how the differences in growth between those groups answers the key question underlying the study: whether poor children given free breakfasts grew faster than those who were not given free breakfasts.*

Magnitude quantifies the **size** of the difference between two numbers.

A hypothesis should state the predicted direction of association between the independent and dependent variables.

Example: Hypothesis: That children from poor families who participate in a free breakfast program will grow faster than those who do not receive a free breakfast.

*The phrase "grow faster" conveys a hypothesized **positive association** between program participation (the **independent variable**) and growth in height (**dependent variable**).*

The most effective descriptions of numeric patterns convey both the direction and size of the pattern. Hypotheses should state expected direction but typically do not predict exact magnitude.

TERMS AND CONCEPTS

Independent variable 20
Risk factor 20
Predictor (variable) 20
Explanatory variable 20
Interpreting numbers 26
Magnitude (size) of association 31
Mixed-methods 18
Negative association 29
Inverse association 31
Number 13
Observation 14
Positive association 28
Direct association 31
Qualitative research methods 18
Quantitative research methods 18
Raw data 14
Reporting numbers 26
Research circle 18
Inductive research 18
Deductive research 18
Descriptive research 18
The W's 25
What 25
When 25
Where 25
Whom/which 25
How 25
Theory 19
Univariate distribution 30
Composition 30
Value 15
Variable 15

HIGHLIGHTS

- A **data set** is composed of information on a consistent set of variables for all **cases** in a sample or population. **Variables** are characteristics that can take on different **values** for different cases in the data set, whereas **constants** are attributes that are the same for all cases in a data set.
- The **research circle** connects inductive (from **data** to **theory**) and **deductive** (from theory to **hypothesis** to data) research strategies. Depending on the research objective, a study can start at any point on the research circle.
- Research can pursue any of four main **goals**: **descriptive**, **exploratory**, **explanatory**, and **evaluative**. The choice of a research goal for a particular research project does not depend on the topic but rather on the question or perspective that is the aim of that research.
- The **W's** include **what** topics or concepts were measured, **where**, **when**, and **from whom** those data were collected, and "**how**" they were collected and analyzed. The W's provide a good mental checklist of the information needed to make sense of the results of a quantitative study.
- For quantitative research documents intended to answer a research question, numbers should be both **reported** and **interpreted**. For research documents intended to provide data mainly for others' use, the numbers may be reported but not interpreted.
- Interpretation of a numeric pattern should describe both the **direction** and **magnitude** of that pattern.

RECOMMENDED READINGS

Ghose, T. (2013, April 2). "Just a theory": 7 misused science words. *Scientific American*. https://www.scientificamerican.com/article/just-a-theory-7-misused-science-words

Utts, J., & Heckard, R. (2014). Turning data into information. *Mind on statistics* (5th ed., pp. 14–67). Cengage, Brooks Cole.

EXERCISES

Individual Exercises

Quantitative Reasoning in Everyday Life

1. Plan a data set about restaurants so that you and your friends have a consistent set of information to help you decide where to eat out. Specify (a) what constitutes a case; (b) a characteristic that is a constant across all cases; and (c) the topics of several variables (things you will measure for each case). (d) Sketch a grid of your data set like that in Figure 2.1, showing (i) what goes in the rows (fill in four or five possible items); (ii) what goes in the columns (fill in four or five possible items); and (iii) possible values for each variable for each case (in the interior cells of the grid).

2. Find a newspaper article or website that compares two or more numeric values on a topic of interest to you. Determine whether the authors described the direction and magnitude of differences between those numbers or left it to readers to do the mathematical comparison.

Identifying and Interpreting Research

3. On the website for the Inter-University Consortium for Political and Social Research (ICPSR; https://www.icpsr.umich.edu), identify a study on a topic of interest to you, either from their list of thematic collections or browse by topic. On the web page for that study, identify (a) the name of the data set and its acronym (e.g., "Civil Rights Data Collection [CRDC]") and (b) the W's (when, where, and from whom data were collected). **Save the URL for the ICPSR website for that data set study to use in exercises for later chapters.** From the tab titled "Data & Documentation," preview or download the codebook or documentation. (c) Use it to identify what constitutes a case in their data set (e.g., type of entity). (d) On the tab titled "Variables" find information on (i) a characteristic that is a constant across all cases in that data set and (ii) the topics of several variables.

4. For each of the following articles or reports, identify (a) its research **goal** (descriptive, exploratory, explanatory, or evaluative) and (b) which research **strategy** it employed (inductive, deductive, or descriptive): Klapper et al. (2015), Chapman et al. (2016), and Crowley (2011). References are in the list of references at the end of the book. Specify the sentences and page numbers in each article that conveyed the information needed to answer each question.

5. Use information from the article by Williams et al. (1997) to identify the following: (a) what constitutes a case in their analysis; (b) the W's (when, where, and to whom the data pertain); (c) the dependent variable in their hypothesis; (d) the main independent variable in their hypothesis; and (e) the hypothesized direction of association between those two variables. For each step, specify the page and paragraph number where you found that information.

6. Repeat the instructions to Exercise 2 but for a journal article or research report.

Planning Research

7. Repeat the instructions to Exercise 1 but for a data set about students' current employment.

Group Exercises

Interpreting Research

8. Discuss the following aspects of the article by Chapman et al. (2016): (a) Identify the process by which they developed their theory or theories (e.g., from what data or other types of information). (b) Write a hypothesis about the relationship implied by their theory, including the expected direction of association. (c) Consider whether they also describe other theories of the same phenomenon. If so, write a hypothesis for each theory.

Planning Research

9. Compare the data sets you designed for Exercise 7 above (about employment), identifying similarities and differences in terms of (a) what constitutes a case in each of your hypothetical data sets; (b) a characteristic that is a constant in that data set; and (c) the topics of several variables. (d) Discuss how these differences could affect your ability to compare information across your data sets.

10. Decide on a social issue of interest. For each of the following types of research goals, discuss possible ways to conduct a study of that topic: (a) descriptive, (b) exploratory, (c) explanatory, and (d) evaluative. For each of those hypothetical projects, discuss whether you would use a descriptive, inductive, or deductive research strategy and why you would choose that strategy.