1 THE DEFINITION AND MEASUREMENT OF CONCEPTS

LEARNING OBJECTIVES

In this chapter, you will learn how to:

1.1 Formulate conceptual definitions that clarify the meaning of research concepts.

1.2 Transform abstract conceptual definitions into operational definitions that can be measured and analyzed effectively.

1.3 Identify and differentiate between systematic and random measurement errors, understanding their impacts on research outcomes.

1.4 Evaluate the reliability and validity of measurements and understand the implications of these criteria for political analysis.

1.5 Recognize the significance of datasets, codebooks, and data analysis software for organizing, analyzing, and interpreting research data.

Think for a moment about the political choices people make. Perhaps most obviously, we vote in elections. But before we vote, we can show our support for a candidate by attending a campaign event, putting up a yard sign, or encouraging friends to vote for our preferred candidate. Our representatives decide which bills they'll sponsor and support. The effects of bills that become laws depend on how they're funded and enforced, whether judges decide to strike them down, whether legislators decide to amend them, not to mention decisions made by businesses, the media, and special interest groups. All these decisions require people to evaluate different options (including the possibility of not deciding) and determine which option they prefer. Politics, after all, is all about making choices.

Our *preferences* help us discuss and describe the world. It is virtually impossible to think about people, places, or things without mentally sorting them according to whether we like them or not and how strongly we like or dislike them. You use your preferences to vote for your preferred candidate on a ballot, decide what to order on a menu, or pick a show to watch on Netflix. Your feelings about things, however, are not tangible and concrete the way the people and things you evaluate are. You cannot see or hear a preference the same way you can see a pro-gun candidate or a gun permit. Preference is a **concept**, an idea or mental construct that organizes, maps, and helps us understand phenomena in the real world and make choices. You can sort and organize objects according to your preferences, mentally separating things you like from things you dislike, then perhaps further separating the things you really like from the things you just like, and so on. Of course, personal preference is not the only criterion for a mental map of the world; for example, you could sort and organize things according to their weight, commercial value, or how politically controversial they are. Some political concepts are quite complicated: *globalization, power, democratization*. Other political concepts, such as *political participation* or *social status*, are somewhat simpler.

PRACTICE THESE SKILLS WITH SOFTWARE

You can practice the skills discussed in this chapter with software of your choice using our *Companions to Political Analysis*.

- **R Companion, 3rd Ed.,** Chapter 1: "Using R for Data Analysis,"
- **Stata Companion, 5th Ed.,** Chapter 1: "Using Stata for Data Analysis,"
- **Excel Companion, 1st Ed.,** Chapter 1: "Using Excel for Data Analysis,"
- **SPSS Companion, 6th Ed.,** Chapter 1: "Introduction to SPSS," and
- **SPSS Companion, 7th Ed.** (forthcoming in 2025/26), Chapter 1: "Using SPSS for Data Analysis."

Whether simple or complicated, concepts are everywhere in political debate, in journalistic analysis, in ordinary discussion, and, of course, in political research. How are concepts used? In partisan or ideological debate—debates about values—concepts can evoke powerful symbols with which people easily identify. A political candidate, for example, might claim that their agenda will ensure "freedom," create "equality," or foster "self-determination" around the globe. These are evocative ideas, and they are meant to be. In political research, concepts are not used to stir up primitive emotional responses. Quite the opposite. In empirical political science, concepts refer to facts, not values. When political researchers discuss ideas like *freedom, equality*, or *self-determination*, they are using these ideas to summarize, label, and understand observable phenomena and tangible things in the real world.

The primary goals of political research are to describe concepts and to analyze the relationships between them. A researcher may want to know, for example, if *social trust* is declining or increasing in the United States, whether political elites are more *tolerant* of dissent than are ordinary citizens, or whether *economic development* causes *democracy*. A **conceptual question**, a question expressed using ideas, is frequently unclear and thus is difficult to answer empirically. A **concrete question**, a question expressed using tangible properties, can be answered empirically. To take a scientific approach to politics, one should try to turn conceptual questions into concrete questions. We don't work on concrete questions because we're not interested in concepts. Nothing could be further from the truth. Because concepts are important, we want to study them productively to better understand the world.

The tasks of describing and analyzing concepts—social trust, political elites, tolerance of dissent, economic development, democracy, and any other concepts that interest us—present formidable obstacles. In her path-breaking book, *The Concept of Representation*, Hanna Pitkin describes the challenge of defining concepts such as representation, power, and interest. She writes that instances "of representation (or of power, or of interest) . . . can be observed, but the observation always presupposes at least a rudimentary conception of what representation (or power, or interest) *is*, what *counts as* representation, where it leaves off and some other phenomenon begins."[1] We need to somehow transform concepts into concrete terms, to express vague ideas in such a way that they can be described and analyzed.

Conceptual definitions are covered in depth in the first section of this chapter. A **conceptual definition** clearly describes the concept's measurable properties and specifies the units of analysis (e.g., people, nations, states, and so on) to which the concept applies. Having clarified and defined a concept, we must then describe an instrument for measuring the concept in the real world. An **operational definition** describes the instrument to be used in measuring the concept and putting a conceptual definition "into operation."

In describing a measurement strategy, we keep an eye trained on the conceptual world: Does this operational definition accurately reflect the meaning of the concept? In this chapter, we consider problems that can emerge when researchers decide on an operational definition. In Chapter 2, we take a closer look at variables, the concrete measurements of concepts.

1.1 CONCEPTUAL DEFINITIONS

As we stated in the chapter introduction, a conceptual definition clearly describes the concept's measurable properties and specifies the units of analysis to which the concept applies. It is important to clearly define concepts because the same concept can and often does mean something different in one context than another or mean different things to different people. Researchers define concepts to make their intended meaning clear to others. If a word or concept means different things to different people, research is likely to be misunderstood.

For example, we could ask you, "Are women more liberal than men? Yes or no?" You might reply, "It depends on what you mean by *liberal*." This is a conceptual question because it uses the intangible term "liberal" and thus does not readily admit to an empirical answer. Are we asking if women are more likely than men to support abortion rights, gun control, government support of education, spending to assist poor people, environmental protection, affirmative action, gay and lesbian rights, funding for drug rehabilitation, or what? Do we mean all these things, some of these things, none of these things, or something else entirely? For some, "liberal" may mean support for gun control. For others, the concept might refer to support for environmental protection. Still others might think the real meaning of liberalism is support for government spending to assist the poor.

Consider, then, the following conceptual definition of *liberalism*: Liberalism is the extent to which individuals express support for increased government spending for social programs. We might be able to improve this definition, but it's a good start. This statement clarifies an abstract political preference, liberalism, by making reference to a measurable attribute—expressing support for government spending on social programs. Someone's preference for liberal policies is abstract and not directly observable, so we focus on what we can observe, like someone's expressing support for government social programs in response to a survey. Notice the words, "the extent to which." This phrase suggests that the concept's measurable attribute—expressing support for government spending—varies across people. Someone who expresses support for government spending is more "liberal" than someone who does not support government spending. It is clear as well that this particular definition is meant to apply to individuals.[2]

The conceptual definition of liberalism we have proposed clarifies what liberalism means to us and suggests a way of measuring it. Without a conceptual definition, we cannot hope to answer the question "Are women more liberal than men?"; having defined the concept of liberalism, the question is now answerable. As you can see, in thinking about concepts and defining them, we keep an eye trained on the empirical world: What are the concrete, measurable characteristics of this concept? The first step in defining a concept is to clarify its empirical meaning.

1.1.1 Clarifying a Concept

To clarify a concept, it is often useful to make an inventory of the concept's concrete properties. After settling on a set of properties that best represent the concept, we write down a definition of the concept. This written definition communicates the subjects to which the concept applies and suggests a measurement strategy. Let's illustrate these steps by working through the example introduced earlier: liberalism.

The properties of a concept must have two characteristics. They must be concrete, and they must vary. The abstract term *liberal* must represent some measurable characteristics of people. After all, when we say that a person or group of people is "liberal," we must have some attributes or characteristics in mind. Someone's liberal preferences may be revealed by the choices they make or other characteristics we can observe about them. Moreover, liberalism varies among people. That is, some people have more (or less) of the measurable attributes or characteristics of liberals than other people do. In clarifying a concept, then, we want to describe characteristics that are concrete and variable. What, exactly, are these characteristics?

The mental exercise of making an inventory of a concept's properties can help you to identify characteristics that are concrete and variable. Think of two cases that are polar opposites with respect to the concept of interest. In this example, we are interested in defining liberalism among individuals, so at one pole, we imagine the stereotypical liberal who has all the tell-tale characteristics of liberalism. At the other pole, we imagine the archetype of conservativism who is the antithesis of liberalism. What images of a perfectly liberal person do you see in your mind's eye? What images of a perfect opposite, an antiliberal or conservative, do you see?[3]

For each case, the liberal and the conservative, we make a list of observable characteristics. In constructing these lists, be open and inclusive. This is a creative, idea-generating exercise so allow yourself to brainstorm even if it means some coloring outside the lines. Here is an example of an inventory of measurable properties you might come up with:

A liberal:

- Has low income
- Is a young person
- Lives in a city

- Favors economic regulations
- Expresses support for government-funded health care and public education
- Attends demonstrations in support of women and immigrants
- Believes free-market capitalism is unfair and causes inequality
- Donates money to liberal causes
- Votes for Democrats
- Watches NBA basketball games, MSNBC News
- Is vegetarian, drives a hybrid car
- Listens to urban music

A conservative:

- Has high income
- Is an older person
- Lives in the suburbs or a rural area
- Favors free-market enterprise
- Expresses opposition to government-funded health care, support for school vouchers
- Attends demonstrations in support of the Tea Party and conservative causes
- Believes free-market capitalism is fair and reduces inequality
- Donates money to conservative causes
- Votes for Republicans
- Watches UFC fights, Fox News
- Plays golf, drives an SUV
- Listens to country music

Brainstorming the measurable properties of a concept is an open-ended process. It produces the raw materials for building conceptual definitions. Once the inventory is made, however, we need to become more critical and discerning. Three problems often arise during the inventory-building process. First, we might think of empirical attributes that are only loosely related to the concept of interest. Second, the inventory may include concepts rather than measurable properties. Third, the empirical properties may represent different dimensions of the concept.

Consider the first three characteristics. According to the list, a liberal "has low income," "is a young person," and "lives in a city," whereas a conservative "has high income," "is an older person," and "lives in the suburbs or a rural area." Think about this for a moment. Are people's income, age, and residence really a part of the concept of liberalism? Put another way: Can we think about what it means to be liberal or conservative without thinking about income, age, and residence? You would probably agree that we could. To be sure, liberalism may be related to demographic factors, such as income, age, and residence, but the concept is itself distinct from these characteristics. This is the first problem to look for when clarifying a concept. Some traits seem to fit with the portraits of the polar-opposite subjects, but they are not essential to the concept. We could say the same thing about what liberals and conservatives tend to watch on television, eat, drive, and do for fun. It's possible we could identify liberals and conservatives based on demographic characteristics and some nonpolitical behaviors, but these things aren't what make someone a liberal or conservative. Let's drop the nonessential traits and reconsider our newly abbreviated inventory:

A liberal:

- Favors economic regulations
- Expresses support for government-funded health care and public education
- Attends demonstrations in support of women and immigrants
- Believes free-market capitalism is unfair and causes inequality
- Donates money to liberal causes
- Votes for Democrats

A conservative:

- Favors free enterprise
- Expresses opposition to government-funded health care, support for school vouchers
- Attends demonstrations in support of the Tea Party and conservative causes
- Believes free-market capitalism is fair and reduces inequality
- Donates money to conservative causes
- Votes for Republicans

After you've brainstormed an inventory of characteristics, imagine that a skeptical observer is looking over your shoulder, pressing you to specify concrete, measurable traits. According to the list, a liberal "favors economic regulations" and "believes free-market capitalism is unfair and causes inequality." A conservative "favors free enterprise" and "believes free-market capitalism is fair and reduces inequality." Neither of these items should be on the list. Why not? Because both terms are themselves abstract concepts. How, exactly, would you determine whether someone supports free enterprise and believes free-market capitalism is fair and can reduce inequality? You can't read their mind or spot these beliefs on a brain-scan image. This is the second problem to look for when clarifying a concept. Some descriptions seem to fit the portraits of the polar-opposite subjects, but these descriptions are themselves vague, conceptual terms that cannot be measured. We should not use one concept to define another; we want to define concepts with concrete, measurable properties. Let's drop the conceptual terms from the inventory.

A liberal:

- Expresses support for government-funded health care and public education
- Attends demonstrations in support of women and immigrants
- Donates money to liberal causes
- Votes for Democrats

A conservative:

- Expresses opposition to government-funded health care, support for school vouchers
- Attends demonstrations in support of the Tea Party and conservative causes
- Donates money to conservative causes
- Votes for Republicans

One could reasonably argue that all these traits belong on an empirical inventory of liberalism. Some observable phenomena that would offer tangible evidence of someone's liberalism, including monetary contributions to issue groups, attending demonstrations, the display of bumper stickers or yard signs, a record of votes cast, or other overt behaviors may be difficult, if not possible, to measure in practice. People have the right to freely associate, vote in secret, and make private contributions to some political organizations, so it may be impossible to know whether someone attended a demonstration, voted for the Democrat or Republican, or gave money to liberal or conservative causes. Depending on the nature of our research and access to data, we may need to focus on characteristics that are readily observed and exclude those that we can't measure.

Examine the remaining inventory items carefully. Can the attributes be grouped into different types? Are some items similar to each other and, as a group, different from other items? A **conceptual dimension** is defined by a set of concrete traits of similar type. You may have already noticed that expressing support for or opposition to government-funded health care and support for public education versus support for school vouchers refer to traditional differences between those who favor a larger public sector and more social services (liberals) and those who favor a more limited governmental role (conservatives). The other items, expressing support for or opposition to gender equality and immigration, refer to more recent disputes between those who favor socially progressive policies (liberals) and those who support traditional social policies (conservatives). This example illustrates the third problem to look for when clarifying a concept. All the traits fit with the portraits of the polar-opposite subjects, but they may describe different dimensions of the concept.

The venerable social science concept of *social status*, for example, has three concrete attributes that vary across people: income, occupation, and education. Yet it seems reasonable to say that all three are empirical manifestations of one dimension of SES.[4] Similarly, if you sought to clarify the concept of *cultural fragmentation*, you might end up with a polar-opposite list of varied but dimensionally similar characteristics of polities: many/few major religions practiced, one/several languages spoken, one/many racial groups, and so on. For each of these concepts, social status and cultural fragmentation, you can arrive at a single measure by determining whether people or polities have a great deal of the concept's characteristics.

Some concepts, such as liberalism, are multidimensional. A **multidimensional concept** has two or more distinct conceptual dimensions. In a multidimensional concept, each conceptual dimension encompasses empirical properties that are similar to each other. Furthermore, each group of traits is qualitatively distinct from other groups of traits. To avoid confusion, the different dimensions need to be identified, labeled, and measured separately. Thus, the traditional dimension of liberalism, often labeled *economic liberalism*, subsumes an array of similar attributes: support for government-funded health care, aid to poor people, funding for education, spending for infrastructure, and so on. The moral dimension, often labeled *social liberalism*, includes policies dealing with gay and lesbian rights, abortion, the legalization of marijuana, the teaching of evolution, and prayer in schools. By grouping similar properties together, the two dimensions can be labeled separately—economic liberalism and social liberalism—and measured separately.[5]

Many ideas in political science are multidimensional concepts. For example, in his seminal work, *Polyarchy*, Robert A. Dahl points to two dimensions of democracy: contestation and inclusiveness.[6] Contestation refers to attributes that describe the competitiveness of political systems—for example, the presence or absence of frequent elections or whether a country has legal guarantees of free speech. Inclusiveness refers to characteristics that measure how many people are allowed to participate, such as the presence or absence of restrictions on the right to vote or conditions on eligibility for public office. Dahl's conceptual analysis has proven to be an influential guide for the empirical study of democracy.[7]

As much as possible, you should define concepts in clear, unidimensional terms. Artists and poets may relish linguistic ambiguity, but social scientists do not. If there are really two separate dimensions of liberalism, we can define and analyze both. Of course, some important political concepts, like power and democracy, are inherently multidimensional, and we should not distort their meaning by attempting to define them in simple, unidimensional terms.

1.1.2 A Template for Writing a Conceptual Definition

After identifying the essential, measurable properties of a concept, we define the concept as clearly as possible. A conceptual definition should communicate three things:

1. The concept being defined,
2. The subjects or groups to which the concept applies, and
3. How the characteristic is to be measured.

The following is a workable template for stating a conceptual definition that meets all three requirements:

The concept of ________ is defined as the extent to which _________ exhibit the characteristic of ____.

For a conceptual definition of economic liberalism, we could write the following:

The concept of <u>economic liberalism</u> is defined as the extent to which <u>individuals</u> exhibit the characteristic of <u>expressing support for government spending for social programs</u>.

Let's consider the template example of a conceptual definition in more detail. The first term, *economic liberalism*, identifies the concept of interest and, when combined with the words "the extent to which," communicates the variation at the heart of the concept. Notice that we're focusing on economic liberalism, as opposed to social liberalism, to avoid conflating two potentially distinct concepts. The second term, *individuals*, states the subjects to whom the concept applies. The third term, *expressing support for government spending for social programs*, suggests how the concept should be measured. Having worked through an inventory of properties of liberalism and thought carefully about what it means, we've identified a concrete and variable characteristic of liberalism that's measurable. This definition of economic liberalism conveys all the essential elements of a conceptual definition.

1.1.3 Why It's Important to Identify the Unit of Analysis

By referring to a subject or group of subjects, a conceptual definition conveys the units of analysis. A **unit of analysis** is the entity we want to describe and analyze. Concepts like preference, conservatism, and representation are inherently abstract; we understand them by studying something we can observe, like a person, city, country, county, university, state, bureaucratic agency, etc. The entity to which the concept applies is the unit of analysis. Students learning the essentials of political analysis sometimes confuse the unit of analysis with the topic of analysis or perhaps the unit of measurement, but it's important to clearly identify the objects of your analysis. The conclusions you draw from analysis depend on the level of your analysis.

Units of analysis can be either individual level or aggregate level. When a concept describes a phenomenon at its lowest possible level, it is using an **individual-level unit of analysis**. Most polling or survey research deals with concepts that apply to individual persons, which are the most common individual-level units of analysis you will encounter. Individual-level units are not always human beings, however. If you were conducting research on the political themes contained in the Democratic and Republican Party platforms over the past several elections, the units of analysis would be the individual platforms from each year. Similarly, if you were interested in finding out whether environmental legislation was a high priority in Congress, you might examine each bill that is introduced as an individual unit of analysis.

Much political science research deals with the **aggregate-level unit of analysis**, which is a collection of individual entities. Neighborhoods or census tracts are aggregate-level units, as are congressional districts, states, and countries. A university administrator who wonders if student satisfaction is affected by class size would gather information on each class, an aggregation of individual students. Someone wanting to know whether states with lenient voter registration laws have higher voter turnout than states with stricter laws could use voter registration laws and voting data from 50 aggregate-level units of analysis, the states. Notice that collections of individual entities, and thus overall aggregate levels, can vary in size. For example, both congressional districts and states are aggregate-level units of analysis—both are collections of individuals within politically defined geographic areas—but states usually represent a higher level of aggregation because they are composed of more individual entities.

There are two general types of aggregate-level data. Some aggregate-level data are really a summary of individual-level units calculated by combining or averaging individual-level characteristics or behaviors, such as an average of student evaluations, the proportion of adults who voted, or some other average characteristic of those in a city, county, or legislative district. Aggregate-level data may also measure the group's characteristics when acting as a group. For example, one could identify which states have lenient voter registration policies and which have strict policies.

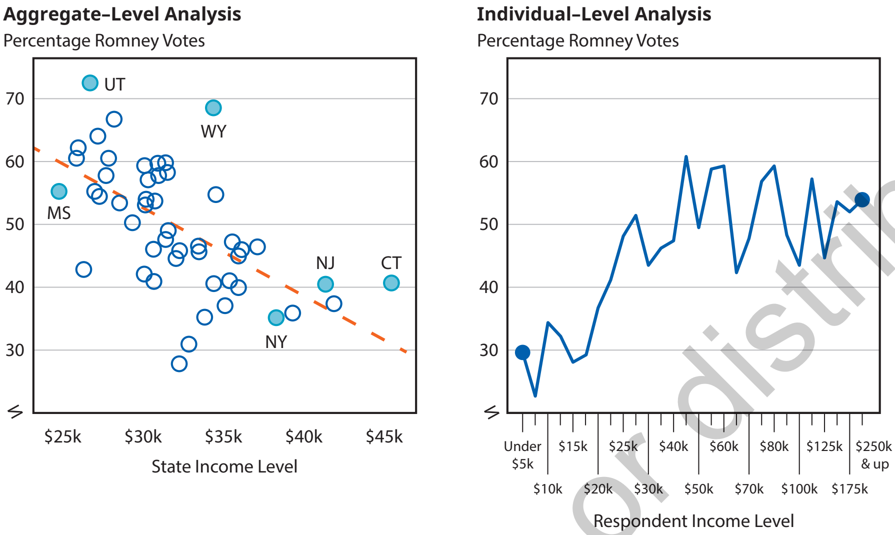
The same concept often can be defined at both the individual and aggregate levels. Ponder this point for a moment. Just as economic liberalism can be defined for individual persons, economic liberalism can be defined for states by aggregating the numbers of state residents who support or oppose government spending: The concept of economic liberalism is defined as the extent to which states exhibit the characteristic of having residents who support government spending for social programs. This conceptual definition makes perfect sense. One can imagine comparing states that have a large percentage of pro-spending residents with states having a lower percentage of pro-spending residents. For statistical reasons, however, the relationship between two aggregate-level concepts usually cannot be used to make inferences about the relationship at the individual level. Suppose we find that states with larger percentages of college-educated people have higher levels of economic liberalism than states with fewer college graduates. Based on this finding, we could not conclude that college-educated individuals are more likely to be economic liberals than are individuals without a college degree.

Sometimes, researchers want to use data collected at one level of analysis to better understand what's happening at another level of analysis. This is called **cross-level analysis**. Cross-level analysis may be necessary when data on certain outcomes are not available at the individual level. For example, a researcher cannot obtain individual-level voting records but may obtain election results by election precinct. Someone interested in juror behavior could compile data on decisions by 6- or 12-member juries but could not observe jury deliberations because they are secret. Researchers interested in health and education outcomes would face similar challenges because of the privacy of medical and educational records.

A classic problem, known as the **ecological fallacy**, may arise when an aggregate-level phenomenon is used to make inferences at the individual level. W. S. Robinson, who coined the term more than 60 years ago, illustrated the ecological fallacy by pointing to a counterintuitive fact: States with higher percentages of foreign-born residents had higher rates of English-language literacy than states with lower percentages of foreign-born residents. At the individual level, Robinson found the opposite pattern, with foreign-born individuals having lower English literacy than native-born individuals.[8] The ecological fallacy is not new, but it continues to create problems and cause confusion.[9] The issue is not that generalizing from one level of analysis to another is always wrong, but sometimes it isn't, and it's difficult to know when it is wrong.[10]

Consider, for example, an aggregate-level analysis of the relationship between income and partisanship in the 2012 presidential election. If one analyzes the relationship between per-capita income and votes for the Republican candidate Mitt Romney with states as the unit of analysis (the left side of Figure 1-1), it appears that poor states are "red states" and rich states are "blue states." It's tempting to infer from this aggregate-level relationship that poor people are more likely to vote Republican than people with higher incomes. Many political pundits read the national electoral map this way, but it's an ecological fallacy. An aggregate-level relationship may not be reflected at the individual level. In fact, an individual-level analysis of the relationship between income and vote choice in the 2012 election shows the *opposite pattern*: as individual income increases, so does the percentage of self-reported Romney voters (the right side of Figure 1-1).

FIGURE 1-1 ■ Illustration of Ecological Fallacy in Vote Choice



A proper conceptual definition needs to specify the units of analysis. Researchers must be careful when drawing conclusions based on the study of aggregate-level units of analysis.

Here's another example of an ecological fallacy that may surprise you. One way to measure the effectiveness of legislators is by calculating their "hit rates," the proportion of their sponsored bills that become laws. We can compare hit rates to identify effective legislators, but we need to be careful about cross-level inferences. Consider this real-life example. One legislator in the Georgia State Assembly, Tim Golden, had a higher hit rate than his colleague, Henry Howard, in the 1999/00, 2001/02, and 2003/04 terms (see Table 1-1). Who had a higher hit rate over all three terms? Howard did. Golden had a higher hit rate than Howard in each term, but Howard had the higher hit rate overall. See for yourself.

TABLE 1-1 ■ Ecological Fallacy in Legislator Hit Rates

Legislator	1999/00 Term	2001/02 Term	2003/03 Term	Overall
Tim Golden	1 / 27 = .037	5 / 15 = .333	8 / 20 = .400	14 / 62 = .226
Henry Howard	0 / 1 = .000	1 / 4 = .250	2 / 8 = .250	3 / 13 = .231

The pattern observed at the term level (Golden > Howard) is reversed when the three terms are aggregated (Howard > Golden). The point is to carefully identify the unit of analysis and beware of making cross-level inferences.

1.2 OPERATIONAL DEFINITIONS

By suggesting how the concept is to be measured, a conceptual definition points the way to a clear operational definition.[11] An operational definition describes explicitly how the concept is to be measured empirically. How could we determine the extent to which people hold opinions that are consistent with economic liberalism? What procedure would produce the truest measure of social liberalism? Suppose we wanted to quantify Dahl's inclusiveness dimension of democracy. We would need to devise a metric that combines the different concrete attributes of inclusiveness. Exactly what form would this metric take? Would it faithfully reflect the conceptual dimension of inclusiveness, or might our measure be flawed in some way? This phase of the measurement process, the step between conceptual definition and operational definition, is often the most difficult to traverse. To help you understand how researchers operationalize abstract concepts, let's consider how researchers might measure preferences and support for liberalism.

The concept of preference is essential to public opinion research, but how can we operationalize this concept? Sometimes, people are asked to compare two or more options and identify their favorite one or rank them in preference order. You can ask people about their past choices. If something is sold in the marketplace, we can discover how much people are willing to pay, or accept as payment, in a transaction. There is usually more than one way to operationalize a concept, but they aren't all equally useful. We often put prices on things to quantify how much they're worth, but many important things aren't bought and sold in a marketplace.

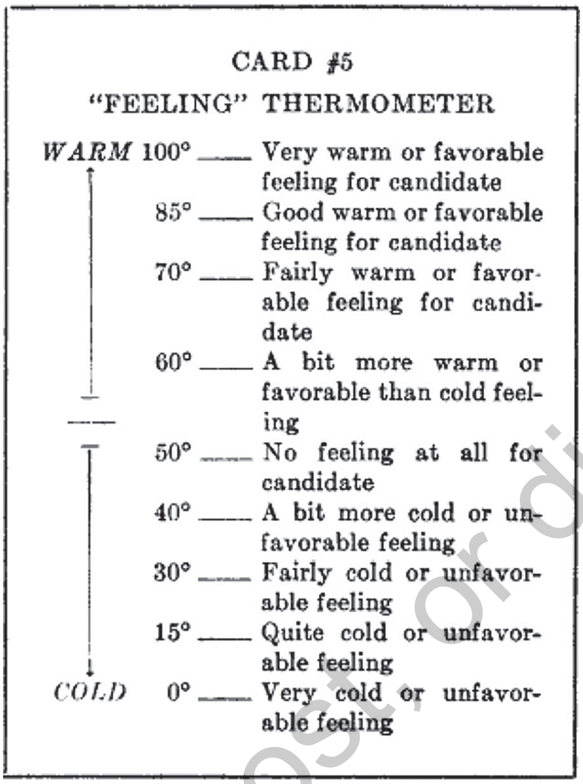
Let's consider a popular method of operationalizing the concept of preference in political science research. Researchers developed a novel method of measuring preferences for the American National Election Study (ANES): the feeling thermometer. A **feeling thermometer** is a visual aid that helps people quantify their feelings about people, ideas, and institutions. It works like this: the researcher shows the respondent a visual aid that calibrates thermometer readings to feelings and asks the following question:

> I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50-degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.

Figure 1-2 shows the card used by ANES interviewers in 1964.[12] As you can see, the feeling thermometer goes from 0 to 100 degrees. Higher numbers correspond to warmer, more favorable feelings, and lower numbers correspond to colder, less favorable feelings. In 1964, this device was used to measure the general public's feelings about presidential candidates, but it's since been broadly deployed to measure the general public's feelings about politicians, groups of people, ideas, and institutions.

Researchers have used feeling thermometers to measure personal preferences for more than 50 years now. Why is the feeling thermometer a good way to operationalize the concept of preference? It's simple and intuitive. People already know how the weather feels. If the temperature is 100 degrees outside, it's a very hot day; if it is 0 degrees, it's a very cold day. Preferences are abstract, but they're frequently associated with our sense of temperature as in getting "cold feet" or having "warm feelings." Feeling thermometers allow people to express their preferences on a relatable scale. (It also makes sense as the percentage you like something from 0 to 100 percent.)

FIGURE 1-2 ■ Feeling Thermometer Used in 1964



CARD #5

"FEELING" THERMOMETER

WARM	100°	Very warm or favorable feeling for candidate
	85°	Good warm or favorable feeling for candidate
	70°	Fairly warm or favorable feeling for candidate
	60°	A bit more warm or favorable than cold feeling
	50°	No feeling at all for candidate
	40°	A bit more cold or unfavorable feeling
	30°	Fairly cold or unfavorable feeling
	15°	Quite cold or unfavorable feeling
COLD	0°	Very cold or unfavorable feeling

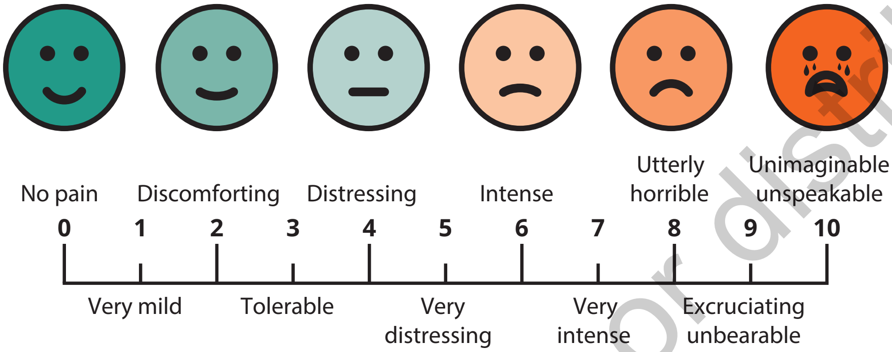
We think feeling thermometers are a good way to measure preferences. Rather than simply take our word for it, try it out for yourself. Reread the block-quoted question prompted earlier and, using Figure 1-2 as a visual aid, rate the following items from the 2020 ANES on a feeling thermometer:

Asian Americans	☐	Feminists	☐	NRA	☐
Blacks	☐	Gay men and lesbians	☐	Planned Parenthood	☐
Black Lives Matter	☐	Hispanics	☐	Police	☐
Big business	☐	Undocumented immigrants	☐	Republican Party	☐
Capitalists	☐	Joe Biden	☐	Scientists	☐
Christians	☐	Journalists	☐	Socialists	☐
Congress	☐	Kamala Harris	☐	Transgender people	☐
Conservatives	☐	Liberals	☐	U.S. Supreme Court	☐
Democratic Party	☐	#MeToo	☐	United Nations	☐
Donald Trump	☐	Muslims	☐	Unions	☐
FBI	☐	NATO	☐		

If you followed the ANES instructions properly, all your ratings should be between 0 and 100. If you don't have positive or negative feelings about an item, you should have scored it 50. Did the feeling thermometer help you quantify your likes and dislikes? What did you do if you did not know the item listed? (In Section 1.4.2, you'll have an opportunity to compare your responses to national averages.)

Recently, physicians have started using a visual aid similar to a feeling thermometer to help people express how much pain they're experiencing. Pain can't be measured directly, but we can picture what it feels like to be in pain. Figure 1-3 shows us how we might operationalize the subjective feeling of pain using a visual aid. If you were asked to quantify the pain you feel from 0 to 10, the faces are really helpful, right?

FIGURE 1-3 ■ Sample Pain Scale



The feeling thermometer was developed to help people quantify their likes and dislikes in face-to-face interviews. It can be used to quantify how much someone likes or dislikes a wide variety of subjects. Of course, no measurement strategy is perfect, and as we'll see, it's always important to evaluate how well we operationalize a concept.

How might we go about implementing the conceptual definition of liberalism? Imagine crafting a series of 10 or 12 survey questions and administering them to many people. Each question would name a specific social program: funding for education, assistance to the poor, spending on medical care, support for childcare subsidies, and so on. For each program, individuals would be asked whether government spending should be decreased, kept the same, or increased. Liberalism could then be operationally defined as the number of times a respondent said "increased." Higher scores would denote more liberal attitudes, and lower scores would denote less liberal attitudes.

As the foregoing examples suggest, an operational definition provides a procedural blueprint for analyzing a concept. An effective measurement strategy unites qualitative and quantitative analysis by allowing researchers to measure abstract concepts. Rather than devalue important concepts like democracy, fairness, and justice, good operational definitions give us the opportunity to better understand and promote these values.

1.3 MEASUREMENT ERROR

Let's use the term *intended characteristic* to refer to the conceptual property we want to measure. The term *unintended characteristic* will refer to any other property or attribute that we do not want our instrument to measure. Given an operational definition, the researcher should ask, "Does this operational instrument measure the intended characteristic? If so, does it measure *only* that characteristic? Or might it also be gauging an unintended characteristic?" Our goal is to devise operational instruments that maximize the congruence or fit between the definition of the concept and the empirical measure of that concept.

Two sorts of error can distort the linkage between a concept and its empirical measure. Serious problems arise when **systematic measurement error** is at work. Systematic error introduces consistent, chronic distortion into an empirical measurement. Often called measurement bias, systematic error produces operational readings that consistently mismeasure the characteristic the researcher is after. Less serious but still troublesome problems occur when **random measurement error** is present. Random error introduces haphazard, chaotic distortion into the measurement process, producing inconsistent operational readings of a concept. To appreciate the difference between these two kinds of error and to see how each affects measurement, we will consider both systematic and random measurement errors in detail. An effective measurement strategy minimizes both systematic and random error, but as we'll see, this ideal is often unachievable, and there may be trade-offs between these two types of measurement error.

1.3.1 Systematic Measurement Error

Suppose that an instructor wants to test the civics knowledge of a group of students. This measurement is operationalized by asking 10 questions about the basic features of American government. First, let's ask, "Does this operational instrument measure the intended characteristic, civics knowledge?" It seems clear that *some* part of the operational measure will capture the intended characteristic, students' actual civics knowledge. But let's press the measurement question a bit further: "Does the instructor's operational instrument measure *only* the intended characteristic, civics knowledge? Or might it also be gauging a characteristic that the instructor did not intend for it to measure?" We know that, quite apart from civics knowledge, students vary in their verbal skills. Some students can read and understand test questions more quickly than others can. Thus, the operational instrument is picking up an unintended characteristic, an attribute it is not supposed to measure—verbal ability.

You can probably think of other characteristics that would "hitch a ride" on the instructor's test measure. In fact, a large class of unintended characteristics is often at work when human subjects are the units of analysis. This phenomenon, dubbed the **Hawthorne effect**, inadvertently measures a subject's response to the knowledge that he or she is being studied. Test anxiety is a well-known example of the Hawthorne effect. Despite their actual grasp of a subject, some students become overly nervous simply because they're being tested, and their exam scores will be systematically depressed by the presence of test anxiety.[13]

The unintended characteristics we have been discussing, verbal ability and test anxiety, are sources of systematic measurement error. Systematic measurement error refers to factors that produce consistently inaccurate measures of a concept. Notice two aspects of systematic measurement error. First, unintended characteristics such as verbal ability and test anxiety are durable, not likely to change very much over time. If the tests were administered again the next day or the following week, the test scores of the same students—those with lower verbal skills or more test anxiety—would yield consistently poor measures of their true civics knowledge. Think of two students, both having the same level of civics knowledge but one having less verbal ability than the other. The instructor's operational instrument will report a persistent difference in civics knowledge between these students when, in fact, no difference exists. Second, this consistent bias is inherent in the measurement instrument. When the instructor constructed a test using word problems, a measure of the unintended characteristic, verbal ability, was built directly into the operational definition. The source of systematic error resides—often unseen by the researcher—in the measurement strategy itself.

A CLOSER LOOK: MEASURING POLITICAL TOLERANCE

Political tolerance is a complex concept, and a large body of research and commentary is devoted to it.[14] Beginning in the 1950s, the earliest research "operationalized" political tolerance by asking large numbers of individuals if certain political freedoms (for example, giving a speech or publishing a book) should be extended to atheists, communists, and socialists. This seemed like a reasonable operational definition because, at the time at least, atheists, communists, and socialists espoused ideas outside the mainstream and were unpopular. The main research finding was somewhat unsettling: Whereas those in positions of political leadership expressed high levels of tolerance, the public at large appeared much less willing to allow basic freedoms for these groups.

Later research, however, pointed to important slippage between the conceptual definition, which clarified and defined the important properties of political tolerance, and the operational definition, the survey questions used to measure political tolerance. The original investigators chose unpopular groups that all had a left-wing or left-leaning ideological bent. The researchers were therefore gauging tolerance only toward leftist groups. Think about this measurement problem. Imagine many people are asked to "suppose that an ardent socialist wants to make a speech in your community. Should an ardent socialist be allowed to speak or not?" For the question's designers, the key words are "wanted to make a speech," and people who respond "allowed to speak" have more political tolerance than those who say "not allowed to speak." But it could be that for some respondents—it is impossible to know how many—the key word is "socialist." These respondents might base their answers on how they feel about socialism, not on their willingness to allow unpopular speech. This operationalization of political tolerance might be measuring respondents' political ideology as much as their political tolerance.

The original survey questions designed to measure political tolerance were ineffective because they measured a characteristic that they were not supposed to measure: individuals' attitudes toward left-wing groups. Thus, the measurement strategy created a poor fit, an inaccurate link, between the concept of tolerance and the empirical measurement of the concept. An effective measurement of political tolerance should accurately gauge individuals' willingness to extend freedoms to unpopular groups.

A better measurement strategy, one more faithful to the concept of political tolerance, allows respondents *themselves* to name the groups they most strongly oppose—that is, the groups most unpopular with or disliked by each person being surveyed. Individuals would then be asked about extending civil liberties to the groups they had identified, not those picked beforehand by the researchers. Think about why this is a superior approach. Imagine people are presented with a list of groups—racists, communists, NRA members, transexuals, capitalists, feminists, and so on—and are asked to name the group they "like the least." Now recast the earlier survey instrument: "Suppose that [a member of the least-liked group] wanted to make a speech in your community. Should he be allowed to speak or not?" Because the respondents selected their least-liked group, investigators can be confident that those who say "allowed to speak" have more tolerance than those who say "not allowed to speak." Interestingly, this superior measurement strategy led to equally unsettling findings: Just about everyone, elites and nonelites alike, expressed rather anemic levels of political tolerance toward the groups they liked the least.[15]

1.3.2 Random Measurement Error

Now, consider some temporary or haphazard factors that might come into play during an instructor's civics knowledge test. Some students may be ill or tired; others may be well rested. Students sitting near the door may be distracted by commotion outside the classroom, whereas those sitting farther away may be unaffected. Commuting students may have been delayed by traffic congestion caused by a fender bender near campus, and so, arriving late, they may be pressed for time. The instructor may make errors in grading the tests, accidentally increasing the scores of some students and decreasing the scores of others.

These sorts of factors—fatigue, commotion, unavoidable distractions—are sources of random measurement error. Random measurement error refers to factors that produce inconsistently inaccurate measures of a concept. Notice two aspects of random measurement error. First, unintended characteristics such as commotion and grading errors are not durable, and they are not consistent across students. They may or may not be present in the same student if the test were administered again the next day or the following week. A student may be ill or delayed by traffic one week, well and on time the next. Second, chance events certainly can affect the operational readings of a concept, but they are not built into the operational definition itself. When the instructor constructed the exam, he did not build traffic accidents into the measure. Rather, these factors intrude from outside the instrument. Chance occurrences introduce haphazard, external "noise" that may temporarily and inconsistently affect the measurement of a concept.

Political scientists who use feeling thermometers to measure public sentiments about political candidates, controversial groups, and ideas also encounter random measurement errors. People taking these surveys have the same issues with fatigue, commotion, and unavoidable distractions that students taking tests do. In addition to these random factors, people will usually round off their reported feeling thermometer scores to a multiple of 5 or 10. So rather than rate their feeling at 73 degrees, they'll say 70 or 75 degrees. The same respondent may round some responses up and other responses down without a clear or consistent pattern of mental accounting, making it a source of random measurement error.

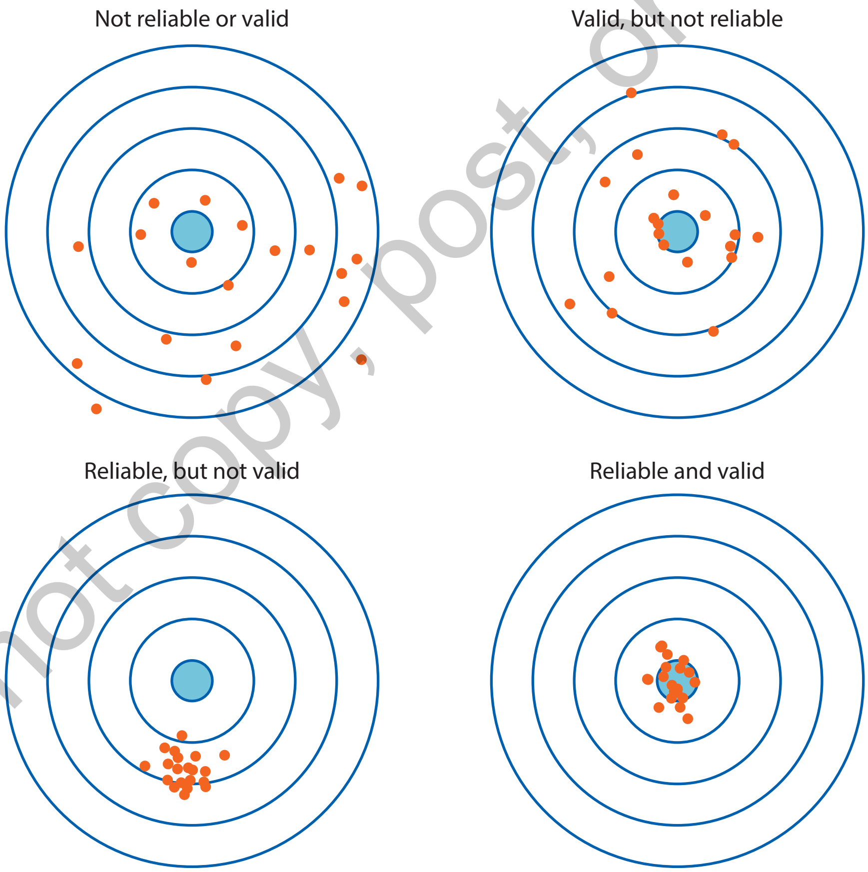
1.4 RELIABILITY AND VALIDITY

We can effectively use the language of measurement error to evaluate the pros and cons of a particular measurement strategy. For example, we could say that the earliest measure of political tolerance, though perhaps having a small amount of random error, contained a large amount of systematic error. The hypothetical instructor's measurement of civics knowledge sounds like it had a dose of both kinds of error—systematic error introduced by durable differences between students in verbal ability and test anxiety and random error that intruded via an array of haphazard occurrences.

Typically, researchers do not evaluate a measure by making direct reference to the amount of systematic error or random error it may contain. Instead, they discuss two criteria of measurement: reliability and validity. However, reliability and validity can be understood in terms of measurement error.

The **reliability** of a measurement is the extent to which it is a consistent measure of a concept. Assuming that the property being measured does not change between measurements, a reliable measure gives the same reading every time it is taken. If multiple researchers are coding information for a study, they're doing it the same way. Applying what we just discussed, a completely reliable measure is one that contains no random error. As random measurement noise increases—repeated measurements jump around haphazardly—a measure becomes less reliable. A measure need not be free of systematic error to be reliable. It just needs to be consistent. If the target centers in Figure 1-4 represent the intended characteristic we want to measure and the points on the targets are our measurement of the characteristic, we assess reliability by the closeness of the marks to one another (regardless of how close they are to the bull's-eye).

FIGURE 1-4 ■ Illustrations of Reliability and Validity



Consider a nonsensical example that nonetheless illustrates the point. Suppose a researcher gauges the degree to which people favor increased government spending on social programs by measuring their blood pressure, with higher pressures denoting stronger approval for spending. This researcher's measure would be reliable. People would have roughly the same blood pressure each time the researcher measured, with some random fluctuation from one day to the next and over the course of the day. But it would clearly be measuring something completely different than opinions about government spending. This poor measurement strategy is represented by the lower-left panel of Figure 1-4. Measuring support for spending using blood pressure readings would be consistent—consistently wrong, that is.

In a more realistic vein, suppose the civics instructor recognized the problems caused by random occurrences and took steps to greatly reduce these sources of random error. Certainly, his measurement of civics knowledge would now be more consistent, more reliable. However, it would not reflect the true civics knowledge of students because it would still contain systematic error. More generally, although reliability is a desirable criterion of measurement—any successful effort to purge a measure of random error is a good thing—it is a weaker criterion than validity.

The **validity** of a measurement is the extent to which it records the true value of the intended characteristic and does not measure any unintended characteristics. A valid measure provides a clear, unobstructed link between a concept and the empirical reading of the concept. Framed in terms of measurement error, the defining feature of a valid measure is that it contains no systematic error, no bias that consistently pulls the measurement off the true value.

To illustrate measurement validity, suppose a researcher gauges opinions toward government spending by asking each respondent to indicate their position on a seven-point scale, from "spending should be increased" on the left to "spending should be decreased" on the right. Is this a valid measure? A measure's validity is harder to establish than is its reliability. But it seems reasonable to say that this measurement instrument is free from systematic error and thus would closely reflect respondents' true opinions on the issue. Or suppose the civics instructor tries to alleviate the sources of systematic error inherent in his test instrument—switching from word problems to an oral examination with visual aids, and perhaps easing anxiety by shortening the test or lengthening the allotted time. These reforms would reduce systematic error, strengthen the connection between true civics knowledge and the measurement of civics knowledge, and thus enhance the validity of the test.

Suppose we have a measurement that contains no systematic error but contains some random error. This situation is represented by the upper-left panel of Figure 1-4. Would this be a valid measure? Can a measurement be valid but not reliable? Although we find conflicting scholarly answers to this question, let's settle on a qualified yes.[16] Instead of considering a measurement as either not valid or valid, think of validity as a continuum, with "not valid" at one end and "valid" at the other. An operational instrument that has serious measurement bias, lots of systematic error, would reside at the "not valid" pole, regardless of the amount of random error it contains. An instrument with no systematic error and no random error would be at the "valid" end. Such a measure would return an accurate reading of the characteristic that the researcher intends to measure, and it would do so with perfect consistency. Now, consider two measures of the same concept, neither of which contains systematic error but one of which contains less random error. Because both measures vanquish measurement bias, both would fall on the "valid" side of the continuum. But the more consistent measure would be closer to the "valid" pole.

1.4.1 Evaluating Reliability

Methods for evaluating reliability are designed around this assumption: If a measurement strategy is reliable, it will yield consistent results. In everyday language, "consistent" generally means "stays the same over time." Accordingly, some approaches to reliability apply this measure-now-measure-again-later intuition. Other methods used to assess the internal consistency of an instrument do not require readings taken at different points in time.

There are several methods of evaluating whether a measurement system is consistent over time. In the **test-retest method**, the investigator applies the measure once and then applies it again at a later time to the same units of analysis. If the measurement is reliable, then the two results should be the same or very similar. If a great deal of random measurement error is present, then the two results will be very different. For example, suppose we construct a 10-item instrument to measure individuals' levels of economic liberalism. We create the scale by asking each respondent whether spending should or should not be increased on 10 government programs. We then add up the number of programs on which the respondent says "increase spending." We administer the questionnaire and then readminister it at a later date to the same people. If the scale is reliable, then each person's score should change very little over time.

The alternative-form method is similar to the test-retest approach. In the **alternative-form method**, the investigator administers two different but equivalent versions of the instrument. The researcher measures the characteristic using one form of the instrument at time point 1 and then measures it again with an equivalent form of the instrument at time point 2. For our economic liberalism example, we would construct two 10-item scales, each of which elicits respondents' opinions on 10 government programs. Why go to the trouble of devising two different scales? The alternative-form method remedies a key weakness of the test-retest method: In the second administration of the same questionnaire, respondents may remember their earlier responses and make sure that they give the same opinions again. Obviously, we want to measure economic liberalism, not memory retention.

Methods for evaluating reliability based on consistency over time have two main drawbacks. First, these approaches make it hard to distinguish random error from true change. Suppose that between the first and second administrations of the survey, a respondent becomes more economically liberal, perhaps scoring a 4 the first time and a 7 the second time. Methods of evaluating reliability over time assume that the attribute of interest—in this case, economic liberalism—does not change over time. Thus, the observed change, from 4 to 7, is assumed to be random error. The longer the time period between questionnaires, the bigger this problem becomes.[17] A second drawback is more practical: Surveys are expensive projects, especially when the researcher wants to administer an instrument to a large number of people.

As a practical matter, most political researchers face the challenge of evaluating the reliability of a measurement that was made at a single point in time. Internal consistency methods are designed for these situations. One internal consistency approach, the **split-half method**, is based on the idea that an operational measurement obtained from half of a scale's items should be the same as the measurement obtained from the other half. In the split-half method, the investigator divides the scale items into two groups, calculates separate scores, and then analyzes the correlation between measurements. If the items are reliably measuring the same concept, then the two sets of scores should be the same.

Following this technique, we would break our 10 government spending questions into two groups of five items each, calculate two scores for each respondent, and then compare the scores. Plainly enough, if we have devised a reliable instrument, then the respondents' scores on one five-item scale should match closely their scores on the other five-item scale.

A more sophisticated internal consistency approach, **Cronbach's alpha**, is a natural methodological extension of the split-half technique. Instead of evaluating consistency between separate halves of a scale, Cronbach's alpha compares consistency between pairs of individual items and provides an overall reading of inter-item correlation and a measure's reliability.[18] Imagine a perfectly consistent measure of economic liberalism. Every respondent who says "increase spending" on one item also says "increase spending" on all the other items, and every respondent who says "do not increase spending" on one item also says "do not increase spending" on every other item. In this scenario, Cronbach's alpha would report a value of 1, denoting perfect reliability. If responses to the items betray no consistency at all—opinions about one government program are not related to opinions about other programs—then Cronbach's alpha would be 0, telling us that the scale is completely unreliable. Of course, most measurements' reliability readings fall between these extremes.

It is easy to see how the methods of evaluating reliability help us to develop and improve our measures of concepts. Let's say we wish to measure the concept of social liberalism, the extent to which individuals accept new moral values and personal freedoms. After building an inventory of this concept's empirical properties, we construct a scale based on support for five policies: gender-affirming medical care, marijuana legalization, abortion rights, stem cell research, and physician-assisted suicide. Our hope is that by summing respondents' five issue positions, we can arrive at a reliable operational reading of social liberalism. With all five items included, the scale has a Cronbach's alpha equal to 0.6. Some tinkering reveals that, by dropping the physician-assisted suicide item, we can increase alpha to 0.7, an encouraging improvement that puts the reliability of our measure near the threshold of acceptability.[19] The larger point to remember is that the work you do at the operational definition stage often helps you to refine the work you did at the concept clarification stage.

1.4.2 Evaluating Validity

The challenge of assessing validity is to identify durable, unintended characteristics that are distorting an operational measure—that is, to identify the sources of systematic measurement error. To be sure, some sources of systematic error, such as verbal skills or test anxiety, are widely recognized, and steps can be taken to ameliorate their effects.[20] In most situations, however, less well-known factors might be affecting validity. In most situations, the true value of the characteristic the researcher wants to measure, represented by the bull's-eye on the targets in Figure 1-4, is unknown (hence, the reason the researcher is attempting to measure it). If you don't know where the intended target is, how do you know how close you came to it?

Consider a measure that surely is familiar to you: standardized academic tests. The SAT, the Law School Admission Test (LSAT), and the Graduate Record Examination (GRE), among others, tend to return consistent results from one administration to the next and are generally correlated with one another. But the debate about such tests does not center on their reliability. It centers, instead, on their validity: Do these exams measure what they are supposed to measure and only what they are supposed to measure?

Critics argue that because many of these tests' questions assume a familiarity with white, middle-class culture, they do not produce valid measurements of aptitudes and skills. Recall again the earliest measurements of political tolerance, which gauged the concept by asking respondents whether basic freedoms should be extended to specific groups: atheists, communists, and socialists. Because several different studies used this operationalization and produced similar findings, the measure was a reliable one. The problem was that a durable unintended characteristic, the respondents' attitudes toward left-wing groups, was "on board" as well, giving a consistent if inaccurate measurement of the concept.

How can researchers identify systematic measurement errors? Researchers tend to evaluate validity using two different criteria: face validity and construct validity. In the **face validity** approach, the investigator uses informed judgment to determine whether an operational procedure is measuring what it is supposed to measure. "On the face of it," the researcher asks, "are there good reasons to think that this measure accurately gauges the intended characteristic?"

Consider, for example, the face validity of feeling thermometer scores recorded in the 2020 American National Election Study. As you can see in Figure 1-5, the national means on these items vary tremendously, with "Scientists" receiving a warm 78.5 mean score and "Socialists" rounding out the ranking with a 37.5 mean feeling thermometer score. On the face of it, do these feeling thermometer scores appear to accurately gauge how the public feels about different people, ideas, and political institutions?

The informed judgment may come from the researcher's own experience as well as careful review of published literature. Do the rankings shown in Figure 1-5 accord with your own experience and whatever research you've conducted on public opinion? Perhaps seeing Donald Trump's pre- and post-election mean feeling thermometer scores at the bottom of the list gives you pause and makes you wonder about partisan bias. However, Joe Biden's pre and post-election scores are also very low, so there doesn't appear to be clear partisan bias.

To assess face validity, the researcher might also compare the inventory of the concept's properties to the operations definition to make sure all of the essential, measurable properties of the concept are included in the measurement technique. Face validity cannot be empirically demonstrated, but a widely accepted measurement strategy is more valid on its face than one with no proven track record.[21]

Let's consider the face validity of a survey question that's been used to measure the concept of political efficacy, the extent to which individuals believe that they can affect government. Feel free to answer this question yourself.

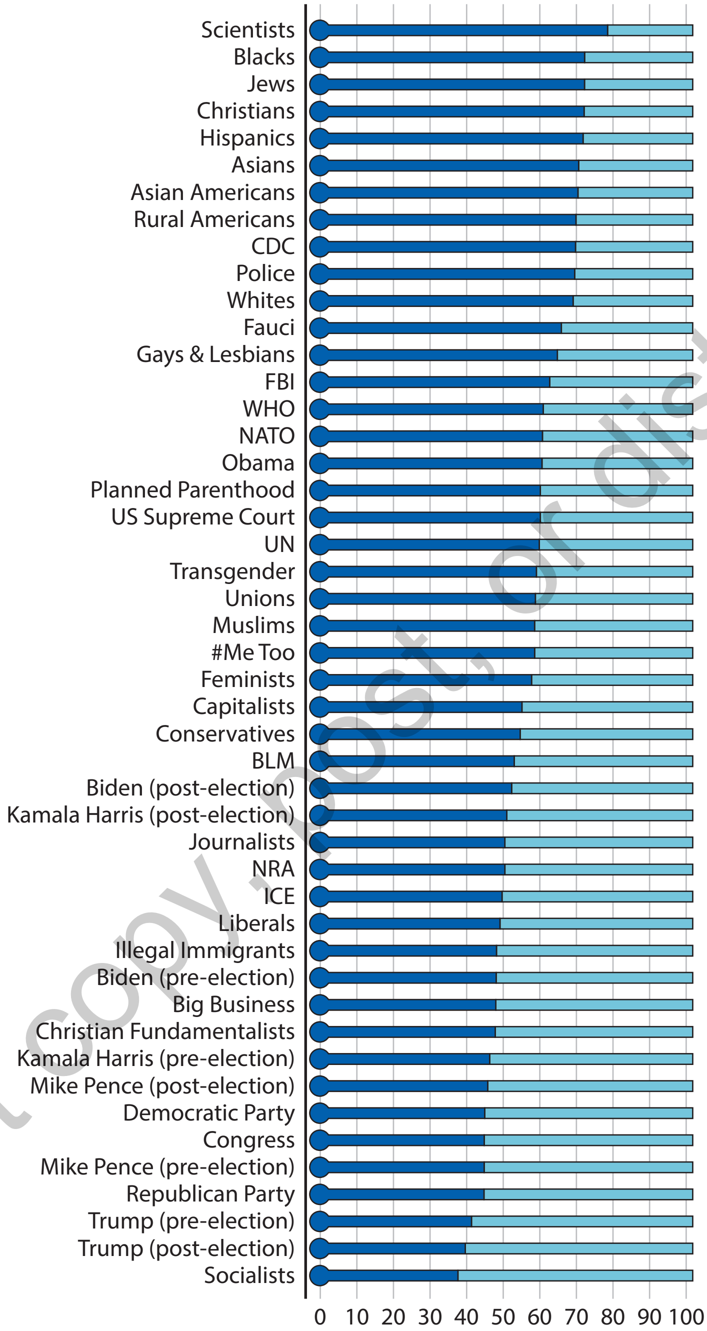
Voting is the only way that people like me can have any say about how the government runs things.

☐ Agree

☐ Disagree

According to the question's operational design, a person with a low level of political efficacy would see few opportunities for influencing government beyond voting and thus would give an "agree" response. A more efficacious person would feel that other avenues exist for "people like me" and so would tend to "disagree." But examine the survey instrument closely. Are there reasons to think that this instrument would not produce an

FIGURE 1-5 ■ National Mean Feeling Thermometer Scores, Highest to Lowest



accurate measurement of the intended characteristic, political efficacy? Consider someone whose sense of efficacy is so weak that they think they have no say in government; to them, voting is not a way to have a say about how the government runs things. At the conceptual level, one would certainly consider such people to have a low amount of the intended characteristic. But how might they respond to the survey question? Quite reasonably, they could say "disagree," a response that would measure them as having a large amount of the intended characteristic. Taken at face value, then, this survey question is not a valid measure.[22] This example underscores a general problem posed by factors that affect validity. We sometimes can identify potential sources of systematic error and suggest how they affect the operational measure. However, it is difficult to know the size of this effect. How many people are being measured inaccurately? A few? Many? It is impossible to know.

On a more hopeful note, survey methodologists have developed effective ways of weakening the chronic distortion of measurement bias, even when the reasons for the bias or its precise size remain unknown. For example, consider the systematic error that can be introduced by the order in which respondents answer a pollster's questions. Consider the following two questions about abortion. Again, feel free to answer them yourself.

Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of serious birth defect in the baby?

☐ Yes

☐ No

Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?

☐ Yes

☐ No

Does the first question affect how you read the second question? It turns out that when the questions are asked in this order, the second question receives a substantially higher percentage of "No" responses than it does otherwise.[23] A solution is available for such question-order effects: Randomize the order in which the questions appear in a survey. In this way, systematic measurement error is transformed into random measurement error. Random measurement error may not be cause for celebration among survey designers, but, as we have seen, random error is easier to deal with than systematic error.[24]

In the **construct validity** approach, the researcher examines the empirical relationships between a measurement and other concepts to which it should be related. Here, the researcher asks, "Does this measurement have relationships with other concepts that one would expect it to have?" For example, if the SAT is a valid measure of high school students' readiness for college, then SAT scores should be strongly related to subsequent grade-point averages earned by college students. If the SAT is an inaccurate measure of readiness, then this relationship will be weak. Evaluating the SAT's construct validity in this manner requires measuring students' academic performance for years after they take the SAT.[25] In applying the construct validity approach, we can use empirical relationships to evaluate an operational measure.

Rest assured that debates about validity in political science are not academic games of "gotcha," with one researcher proposing an operational measure and another researcher marshaling empirical evidence to shoot it down. Rather, the debate is productive. It is centered on identifying potential sources of systematic error, and it is aimed at improving the quality of widely used operational measures. It bears emphasizing as well that although the problem of validity is a concern for the entire enterprise of political analysis, some research is more prone to it than others. A student of state politics could obtain a valid measure of the concept of state-supported education fairly directly by calculating a state's per-capita spending on education. A congressional scholar would validly measure the concept of party cohesion by figuring out, across a series of votes, the percentage of times a majority of Democrats opposed a majority of Republicans. In these examples, the connection between the concept and its operational definition is direct and easy to recognize. By contrast, researchers interested in individual-level surveys of mass opinion, as the examples illustrate, often face tougher questions of validity.

1.5 WORKING WITH DATASETS, CODEBOOKS, AND SOFTWARE

We have already discussed how political science concepts are defined and measured. Conceptual definitions emphasize measurable properties that vary. Operational definitions specify what instruments will be used to measure the concept's empirical properties. An effective measurement strategy produces reliable and valid measures of what the researcher intended to measure. Given all that's required to define and measure concepts properly, it's important to organize the information we generate so it can be analyzed and understood. In this section, we introduce some essential terms and concepts related to this aspect of the research process.

We call the information we collect **data** and organize our data into **datasets**. To be grammatically correct, a singular bit of information is datum (a singular noun) and many bits of datum together are data (a plural noun). "Data are" may sound odd to you, but it's grammatically correct. Kellstedt and Whitten offer their marching orders: "Get used to it: You are now one of the foot soldiers in the crusade to get people to use this word appropriately. It will be a long and uphill battle."[26]

Datasets can be enormous or tiny; they can contain names, dates, large numbers, small numbers, website links, or whatever other information the creator thought to save. Despite enormous variety in content, datasets tend to share the same general structure. When you open a dataset using statistical software, like SPSS, Stata, or R, or other software that allows you to view a dataset, it looks a lot like a spreadsheet with rows and columns (in fact, some datasets are spreadsheets). Each unit of analysis or observation fills a row of the dataset. Each row of a public opinion dataset represents a person who answered the survey. Identification numbers that uniquely identify each row typically fill the dataset's first column, but this is only customary and not required. Each column of the dataset stores the values of a variable. Figure 1-6 shows the beginning of a dataset on roll-call voting in the House of Representatives in the 73rd Congress compiled by Keith Poole and Howard Rosenthal.

Each row of Figure 1-6 represents one U.S. Representative who cast roll-call votes in this historic legislative session. They are uniquely identified by the "id" variable that defines the second column. Each column records values of a variable; a few of these values are text, but most are numbers. Figure 1-6 displays only the first 13 rows and 11 columns of the dataset, which has 450 rows and 152 columns in all.

It's easy to tell what some of the entries shown in Figure 1-6 mean; "cong" is the term for Congress, and "name" is the member's last name. But the meaning of some of these variables isn't self-evident. If you're using a dataset, it's important to know how the authors measured concepts of interest. You can look up variable names, descriptions, and other

FIGURE 1-6 ■ Example of a Dataset on Roll-Call Voting in Congress

	cong	id	state	dist	lstate	party	eh1	eh2	name	V1	V2
1	73	12	47	3	NORTH C	100	0	1	ABERNETHY	1	6
2	73	19	21	15	ILLINOI	100	0	1	ADAIR	1	6
3	73	43	11	1	DELAWAR	100	0	1	ADAMS	1	6
4	73	121	21	13	ILLINOI	200	0	1	ALLEN	2	1
5	73	137	41	5	ALABAMA	100	0	1	ALLGOOD	1	6
6	73	143	41	8	ALABAMA	100	1	1	ALMON	1	6
7	73	189	3	6	MASSACH	200	0	1	ANDREW	2	1
8	73	200	13	40	NEW YOR	200	0	1	ANDREWS	2	1
9	73	227	33	98	MINNESO	537	0	1	ARENS	5	6
10	73	252	21	23	ILLINOI	100	0	1	ARNOLD	1	6
11	73	292	12	14	NEW JER	100	0	1	AUF DER HEI	1	6
12	73	307	64	2	MONTANA	100	0	1	AYERS	1	6
13	73	309	32	5	KANSAS	100	0	1	AYRES	1	6

important information about a dataset in a **codebook**. The codebook for this dataset, for example, informs us that the values in column 3 ("state") refer to two-digit Inter-university Consortium for Political and Social Research (ICPSR) state codes and provides a key to the numeric party codes in the sixth column (100 is the code for Democrats who controlled the House in 1932).[27] We can also find more information about the roll-call votes taken in this Congress (you can see V1 and V2 on the far right of Figure 1-6). The first vote recorded in this Congress, "V1," elected Rep. Henry Rainey, D-IL, to Speaker of the House on March 9, 1933.

If you compile a dataset through original research or create new variables by transforming variables in an existing dataset, document your work carefully so it's clear what you have done. If your dataset is for personal use, you don't need to create a publication-quality codebook, but you should take notes that you can refer to later.

Researchers clearly define concepts and measurement strategies so others can evaluate, replicate, and improve upon their work. Scientific knowledge is transmissible; the knowledge we produce contributes to an ongoing conversation among academic researchers. This is how we build upon prior research and make scientific progress. The data you see recorded in Figure 1-6, for example, have been made available to generations of American politics scholars. Researchers can use this dataset along with datasets on other terms of Congress (from the first term of Congress to the present day). Researchers can also use the identification codes to merge this dataset with additional data on members of Congress and the states they represent.[28]

As you've learned, there are different ways to measure a conceptual property that varies. The property or characteristic that interests us may vary across units of analysis at a given time and it also may vary within the units of analysis over time. A dataset that compiles information collected at one time to study properties that vary across the units of analysis is a **cross-sectional dataset**. Data from cross-sectional studies are the norm in social science research. Most public opinion studies are cross-sections of the population. A **cross-sectional study** contains information on units of analysis measured at one point in time. Respondents a, b, and c are interviewed—that's it.

A dataset that compiles information collected at different time intervals to study properties that vary over time is a **time-series dataset**. Time-series datasets typically record an aggregate-level variable's values at regular time intervals. For example, the president's public approval ratings vary over time and can be measured at regular intervals.

Another type of dataset, called pooled datasets or time-series cross-sectional datasets, incorporates cross-sectional and longitudinal variation. A pooled dataset on public opinion on issues 1, 2, and 3, for example, might ask Respondents a, b, and c questions 1, 2, and 3 one year and ask Respondents x, y, and z questions 1, 2, and 3 the next year. Notice that the pooled dataset asked the same questions to different respondents in years 1 and 2. A special subset of pooled data, panel dataset or panel studies, feature both cross-section and temporal variation by using the same subjects over time. The test-retest and alternative-form approaches to evaluating reliability, discussed earlier, require data obtained from panel studies. A **panel study** contains information on the same units of analysis measured at two or more points in time. Respondents a, b, and c are interviewed at time 1; Respondents a, b, and c are interviewed again at time 2. Panel studies allow researchers to observe variation within each unit, but they're rare gems because researchers must invest significant time and resources to produce them.

SUMMARY

In this chapter, we introduced the essential features of concepts and measurement. A concept is an idea, an abstract mental image that cannot be analyzed until its concrete properties are measured. A main goal of social research is to express concepts in concrete language, to identify the empirical properties of concepts so that they can be analyzed and understood. This chapter described a heuristic that may help you to clarify the concrete properties of a concept: Think of polar-opposite subjects, one of whom has a great deal of the concept's properties and the other of whom has none of the properties. The properties you specify should not themselves be concepts, and they should not describe the characteristics of a different concept. It may be, as well, that the concept you are interested in has more than one dimension.

This chapter described how to write a conceptual definition, a statement that communicates variation within a characteristic, the units of analysis to which the concept applies, and how the concept is to be measured. Important problems can arise when we measure a concept's empirical properties—when we put the conceptual definition into operation. Our measurement strategy may be accompanied by a large amount of random measurement error, error that produces inconsistently incorrect measures of a concept. Random error undermines the reliability of the measurements we make. Our measurement strategy may contain systematic measurement error, which produces consistently incorrect measures of a concept. Systematic error undermines the validity of our measurements. Although measurement problems are a persistent worry for social scientists, all is not lost. Researchers have devised productive approaches to enhancing the reliability and validity of their measures.

KEY TERMS

aggregate-level unit of analysis
alternative-form method
codebook
concept
conceptual definition
conceptual dimension
conceptual question
concrete question
construct validity
Cronbach's alpha
cross-level analysis
cross-sectional dataset
cross-sectional study
data

dataset
ecological fallacy
face validity
feeling thermometer
Hawthorne effect
individual-level unit of analysis
multidimensional concept
operational definition
panel study
random measurement error
reliability
split-half method
systematic measurement error
test-retest method
time-series dataset
unit of analysis
validity

EXERCISES

1. Suppose you wanted to study the role of religious belief, or religiosity, in politics and society. You would begin by setting up an inventoryThrough of empirical properties, contrasting the mental images of a religious person and a nonreligious person.

A religious person:	A nonreligious person:
a. Regularly prays	a. Never prays
b.	b.
c.	c.

A. Item a, "regularly prays/never prays," provides a good beginning for the inventory. Think up and write down two additional items, b and c.

B. As discussed in this chapter, a common problem in developing an empirical inventory is that we often come up with items that measure a completely different concept. For example, in constructing the liberal–conservative inventory, we saw that "has low income"/"has high income" did not belong on the list because income and ideology are different concepts. For each item you chose in part A, explain why you think each property is a measure of religiosity and does not measure any other concept.

C. Using one of your items, b or c, write a conceptual definition of religiosity. In writing the conceptual definition, be sure to use the template presented in this chapter.

2. *Finding 1:* An examination of state-level data on electoral turnout reveals that as states' percentages of low-income citizens increase, turnout increases. *Conclusion:* Low-income citizens are more likely to vote than are high-income citizens.

A. For the purposes of this exercise, assume that Finding 1 is correct—that is, assume that Finding 1 describes the data accurately. Is the conclusion supported? Making specific reference to a problem discussed in this chapter, explain your answer.

B. Suppose that, using individual-level data, you compared the voting behavior of low-income citizens and high-income citizens. *Finding 2:* Low-income citizens are less likely to vote than high-income citizens. Explain how Finding 1 and Finding 2 can both be correct.

3. This chapter discussed the Hawthorne effect, a measurement problem that can arise when people are aware they are being studied. In public opinion surveys, similar measurement issues, *social desirability effects*, can distort expressed levels of support for controversial social policies, such as affirmative action programs that give hiring preferences to Black people. As you can imagine, this problem is often heightened when respondents are aware of the demographic characteristics of the interviewer, such as the interviewer's race or sex. Consider an example using respondents' knowledge of the interviewer's sex. The 2012 General Social Survey asked respondents the following question:

"Do you happen to have in your home (or garage) any guns or revolvers?"

- ☐ Yes
- ☐ No
- ☐ Refused

A. Perform a mental experiment. Visualize a group of respondents, all of whom do, in fact, have guns in their homes. (i) Do you think that a sizeable number of these respondents would be less willing to answer truthfully "yes" if the interviewer were female than if the interviewer were male? (ii) Explain the reasoning behind your answer in (i). (There is no correct or incorrect answer. Just think about it and explain your logic.)

B. Now, think about the two types of measurement error we discussed in this chapter: systematic measurement error and random measurement error. With that difference in mind, suppose you discovered that respondents in the 2012 GSS were substantially less likely to answer "yes" to female interviewers than to male interviewers. (i) Would this be a problem of systematic measurement error or random measurement error? (ii) Explain your answer in (i) in part B, making reference to the difference between the two types of error.[29]

4. Four researchers, Warren, Xavier, Yolanda, and Zelda, have devised different operational measures for gauging individuals' levels of political knowledge. Each researcher's operational measure is a scale ranging from 0 (low knowledge) to 100 (high knowledge). For the purposes of this exercise, assume that you know—but the researchers do not know—that the "true" level of knowledge of a test respondent is equal to 50. The researchers measure the respondent four times. Here are the measurements obtained by each of the four researchers:

Warren: 40, 60, 70, 45
Xavier: 48, 48, 50, 54
Yolanda: 49, 50, 51, 50
Zelda: 45, 44, 44, 46

A. Which researcher's operational measure has high validity and high reliability? Explain.

B. Which researcher's operational measure has high validity and low reliability? Explain.

C. Which researcher's measure has low validity and high reliability? Explain.

D. Which researcher's measure has low validity and low reliability? Explain.

5. Two candidates are running against each other for a seat on the city commission. You would like to obtain a valid measurement of which candidate has more

pre-election support among the residents of your neighborhood. Your operational measure: Obtain a precise count of yard signs supporting each candidate. The candidate with a greater number of yard signs will be measured as having greater pre-election support than the candidate having fewer yard signs.
Recall this chapter's discussion of *face validity*. In assessing face validity, the researcher asks, "Are there good reasons to think that this measure is not an accurate gauge of the intended characteristic?" Clearly, the yard-sign measurement strategy has low face validity, because it clearly measures unintended characteristics—characteristics other than pre-election support for the two candidates. For example, because yard signs cost money, a yard-sign count may be measuring the size of candidates' campaign budgets, not necessarily potential support among the voting public. Describe two additional unintended characteristics that, plausibly, are being measured by a count of the number of yard signs.

6. Mutt Jeffley wants to weigh his dog. He proceeds as follows: While holding the dog, he steps onto a bathroom scale and records the weight. Just to make sure he got it right, he repeats the procedure: While holding the dog, he steps onto the scale a second time and again records the weight. Obviously, Mutt's strategy will produce a faulty measurement of the intended characteristic, the weight of his dog. Review this chapter's discussion of reliability and validity. Again examine Figure 1-4, which uses a target analogy to illustrate combinations of the criteria of measurement.

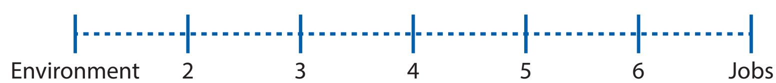
A. Which of these scenarios best fits Mutt's measurement of his dog's weight?

☐ Not reliable or valid
☐ Valid but not reliable
☐ Reliable but not valid

B. Making reference to the characteristics of reliability and validity, explain your answer in A.

C. Describe how Mutt could change his measurement procedure to produce a measurement of his dog's weight that is both valid and reliable.

7. Conflicts that arise in environmental policy are often framed as trade-offs between protecting the environment and creating jobs. The ongoing debate over the Keystone XL Pipeline, which pits environmental groups against the fossil fuels industry, is one example. The spotted owl controversy of the 1990s, which arrayed animal rights activists and environmentalists against logging interests, is another. Survey researchers have sought to measure individuals' opinions on trade-offs such as these. In the traditional measure of the trade-off, respondents are shown a seven-point scale and asked to place themselves at one of the seven positions, from "protect environment, even if it costs jobs and standard of living" at point 1 to "jobs and standard of living more important than environment" at point 7.



A. Think about the *face validity* of this survey instrument. Recall that in evaluating face validity, the researcher asks, "Are there good reasons to think that this measure is not an accurate gauge of the intended characteristic?" In considering its face validity, you may even wish to assess whether this scale would validly measure your own position on the environment-versus-jobs trade-off. (i) Do you think that this scale has high face validity or low face validity? (ii) Explain your answer in (i).

B. Suppose you use this measure in your own survey, obtaining data on a large number of individuals. Suppose further that you decide to test the *construct validity* of the scale. Recall that in evaluating construct validity, the researcher asks, "Does this measurement have relationships with other concepts that one would expect it to have?" For example, researchers have evaluated the construct validity of the party identification scale by seeing how well it relates to voting turnout in primary and general elections: stronger partisans should have higher turnouts than weaker partisans. Consider three possible ways to test the constructive validity of the environment–jobs trade-off scale. One could examine the relationship between the scale and respondents' opinions on (i) abortion, (ii) climate change, or (iii) business regulation. Which one of these three relationships would provide the *best test* of construct validity? Explain your answer. If the scale had high construct validity, what would the relationship "look like"?

8. This chapter discussed the different ways that data are collected and organized for analysis. Of particular importance is the difference between *cross-sectional data* and *longitudinal data*. For each of the situations described in parts A and B, answer the following: (i) State whether the researcher's dataset will be cross-sectional or longitudinal. (ii) Explain how you know.

A. Using data obtained from Freedom House on the 100 largest countries of the world, a researcher plans to analyze the spread of the Internet between 1990 and the present day.

B. Another researcher, using data on the 100 largest countries for the most current year, seeks to analyze the relationship between countries' gross domestic product (GDP) per capita and level of civil unrest.

9. Over the past several years, the term "polarization" has been receiving a lot of attention from political journalists and academics, particularly with regard to American politics. Democratic and Republican voters are said to be "polarized," as are members of the House and Senate. Think for a moment about the concept of *polarization*. To say that the electorate, for example, is polarized, one must also have an idea of what shape a nonpolarized electorate would take. Political scientists have, of course, addressed the measurement issues associated with this concept.

A. Using available Internet resources, such as your library's access to online journals, search for one of the following: *American Journal of Political Science, Journal of Politics*, or *American Political Science Review*. Having located one of these journals, type "polarization" in the search bar and find a scholarly article on the topic. Write down the article's reference: author(s), title, journal, and date. (*Note*: you will need to gain access to the full article, not simply the article's abstract.)

B. Browse the article you cited in part A. Write a paragraph that describes the operational definition of polarization. That is, how is the concept operationally measured in the research article?

2 MEASURING AND DESCRIBING VARIABLES

LEARNING OBJECTIVES

In this chapter, you will learn how to:

2.1 Recognize the essential features of a variable.

2.2 Differentiate between nominal, ordinal, and interval levels of measurement, understanding their implications for data analysis.

2.3 Calculate and interpret measures of central tendency and measures of dispersion, such as range and variance.

2.4 Develop strategies to effectively describe and analyze nominal-level variables.

2.5 Apply appropriate methods to describe and analyze ordinal-level variables.

2.6 Utilize appropriate methods to analyze interval-level variables to describe their distribution and variation.

The operational definition of a concept provides a blueprint for its measurement. When we follow through on the plan, when we construct what the blueprint describes, we end up with a variable. A **variable** is an empirical measurement of a characteristic. Variables provide the raw materials for describing and analyzing the social and political world.

All variables share certain features. After a preliminary dissection of an exemplar variable, we turn to a discussion of levels of measurement, the amount of information conveyed by a variable's values and codes. Some characteristics, such as a person's age, can be measured with greater precision than others, such as a person's marital status. Accordingly, the values and numeric codes of some variables contain more information than do the values and codes of other variables. A variable's level of measurement determines how precisely we can describe it.

PRACTICE THESE SKILLS WITH SOFTWARE

You can practice the skills discussed in this chapter with software of your choice using our *Companions to Political Analysis*. See Chapter 2: "Descriptive Statistics," in ***R Companion, 3rd Edition; Stata Companion, 5th Edition; Excel Companion, 1st Edition, SPSS Companion, 6th Edition, and SPSS Companion, 7th Edition*** (forthcoming in 2025/26).

In this chapter, we also consider the two cornerstones of description: central tendency and dispersion. Central tendency refers to the typical or "average" value of a variable. Dispersion refers to the amount of variation or "spread" in a variable's values. You will find that central tendency and dispersion are not separate aspects of a variable. They work together to provide a complete description of a variable. The final section of this chapter discusses a variety of interesting techniques that political scientists use to transform variables.

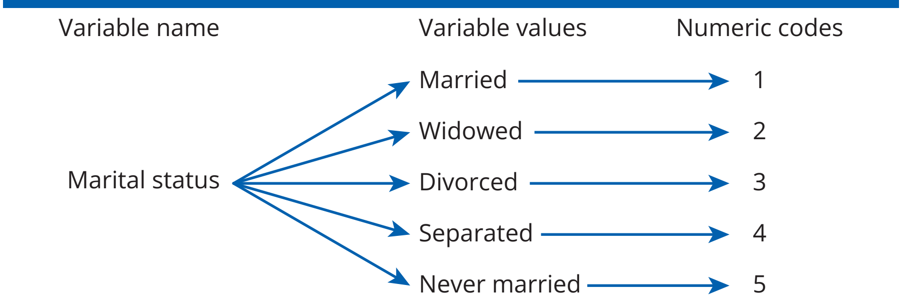
2.1 ESSENTIAL FEATURES

Every variable has one name and at least two values. If there aren't at least two different values, the characteristic being measured doesn't vary; it's a constant, not a variable. Furthermore, if computer analysis is to be performed, then a variable's values must be coded numerically. So a name, two or more values, and numeric codes—these are the gross anatomy of any variable.

For example, many public opinion surveys measure marital status by asking each respondent to choose the category that describes him or her: married, widowed, divorced, separated, or never married. These five categories are the values of the variable, marital status. A person who responds "married" is measured as having a different value on the variable than someone who says "divorced." Similarly, when an application form requests "Age: ___________," it is asking for a value of the variable, age. Someone who writes "20" has a value on the variable that is 9 measurement units (years) younger than someone who writes "29."

Figure 2-1 displays the key features of a variable and introduces essential terminology. It may also help clear up confusion about variables. Like all variables, marital status has one name, and it has at least two values—in this case, five: married, widowed, divorced, separated, and never married. As Figure 2-1 illustrates, the descriptors "married" and "widowed" are different values of the marital status variable, not different variables.

FIGURE 2-1 ■ Anatomy of a Variable



It is not uncommon to get confused about the distinction between a variable's name and its values. They are different values of the same variable, marital status. Here is a heuristic that will help you become comfortable with the distinction between a variable's name and values. Think of one unit of analysis and ask this question, "What is this unit's __________?" The word that fills in the blank is never a value. It is always a variable's name. It makes no sense to complete the question this way ("What is this person's divorced?"). But the question "What is this person's marital status?" makes perfect sense. Ask the question, fill in the blank, and you have the name of a variable—in this case, marital status. Answer the question "What is this person's marital status?" and you have one of the variable's values: "divorced" (or "married," "widowed," "separated," "never married"). The value, "divorced," is one of the values of the variable named marital status.[1]

2.2 LEVELS OF MEASUREMENT

Beyond the fundamental requirement of having at least two values, variables can differ in how precisely they measure a characteristic. Researchers distinguish three levels of precision with which a variable measures an empirical characteristic. It's important to be able to identify a variable's level of measurement because the methods we use to analyze a variable depend on its level of measurement.

The values of the marital status variable enable us to place people into different categories—and nothing more. Furthermore, the numeric codes associated with each value of marital status, code 1 through code 5, merely stand for the different categories—and nothing more. "Married" is different from "never married," and "1" is different from "5." Marital status is an example of a nominal-level variable, a variable whose values and codes only distinguish different categories of a characteristic.

Now, imagine what the anatomy of a different variable, age, would look like. Just as with marital status, age would have one name. And age would have an array of different values, from "18" to (say) "99." Now, move from values to codes. What would be the numeric codes for age? The numeric codes for age would be identical to the values of age. Age is an example of an interval-level variable, which has values that tell us the exact quantity of a characteristic. The values of an interval-level variable convey more information than do the values of a nominal-level variable. Another type of variable, an ordinal-level variable, conveys more information than does a nominal variable but less information than an interval variable. Let's take a closer look at these levels of measurement.

2.2.1 Nominal-Level Variables

Nominal variables are the least precise. A **nominal-level variable** communicates differences between units of analysis on the characteristic being measured. The values of a nominal-level variable are mutually exclusive and completely exhaustive so that each observation fits in just one category. Some nominal-level variables are binary (with only two dichotomous values), while others separate units of analysis into multiple categories. As discussed earlier, marital status is a nominal variable. Its values allow us to separate people into different categories.

The values a nominal variable tell us that subjects having one value, such as Protestant, differ from subjects that have another value, such as Catholic. Gender (female/male), race (white/Black), country of origin (born in the United States/not born in the United States), union membership (member/not a member), employment sector (government employee/private-sector employee)—all these are examples of characteristics that are measured by nominal variables. In each case, the values only represent different categories of the measured characteristic.

As with most variables, the values of nominal variables are frequently recorded using numeric codes. It is important to remember, however, that these codes do not represent quantities. They merely record differences. In most cases, the numeric codes associated with the values of a nominal-level variable are arbitrary. Thus, we might measure religious denomination by five values: Protestant, Catholic, Jewish, other religion, and no religious affiliation. For convenience, we could choose numeric codes 1, 2, 3, 4, and 5 to represent those categories. But we could just as easily choose 27, 9, 56, 12, and 77. The numeric codes themselves do not have inherent meaning. They derive their function from the simple fact that they are different.

For nominal variables, the link between values and numbers can create confusion. Students sometimes mistakenly equate the word *nominal* with *number*. Given this mistaken assumption, for example, one might misidentify an interval-level variable such as age—measured by a number, number of years—as a nominal variable. This would be incorrect. The word *nominal* means "in name only." In French, "nom" means name. Thus, variables whose values are names, or whose numeric codes only represent names, are nominal variables.

2.2.2 Ordinal-Level Variables

Ordinal variables are more precise than nominal-level variables. An **ordinal-level variable** communicates relative differences between units of analysis. Ordinal variables have values that can be ranked. Plus the ranking is reflected in the variable's numeric codes.

Consider an ordinal-level variable named "social trust," which asks how much respondents think people can be trusted: always, most of the time, about half of the time, some of the time, or never. Notice that, just as with nominal variables, the values permit you to classify respondents into different categories. A person who says people can "always" be trusted would be measured as being different from a person who says people can "never" be trusted. Unlike nominal variables, however, the values of an ordinal-level variable permit you to distinguish the *relative amount* of the measured characteristic. Someone who says people can "always" be trusted has a higher level of social trust than someone who says "most of the time," and that person has more social trust than someone who says people can be trusted "about half the time." The values of ordinal variables have numeric codes that reflect the relative amounts of the characteristic. For convenience and simplicity, "always" could be coded 1, "most of the time" 2, "about half the time" 3, "some of the time" 4, and "never" 5. These codes impart the ranking that underlies the values, with 1 being most trusting and 5 being least trusting. But the numeric coding could be reversed with "always" coded 5 and never coded "1," and they would still be in rank order. The numeric codes could go from 6 to 10 or 10 to 6; as long as they're in sequential order, they convey ordinal-level information.

Ordinal-level variables abound in social research, especially survey research. Survey researchers and demographers are interested in measuring geographic mobility, the extent to which people have moved from place to place during their lives. What values are used to measure this variable? Typically, respondents are asked this question: "Do you currently live in the same city that you lived in when you were 16 years old? Do you live in the same state but a different city? Or do you live in a different state?" So the values are "same city," "same state but different city," and "different state." Look at these values and follow the steps. Do the values tell you the exact amount of geographic mobility, the characteristic being measured? No, the values are not expressed in an interval unit, such as miles. So this is not an interval-level variable. Do the values allow you to say that one person has more of the measured characteristic than another person? Can you say, for example, that someone who still lives in the same city has more or less of the characteristic, geographic mobility, than someone who now lives in the same state but in a different city? Yes, the second person has been more geographically mobile than the first. Because the values permit us to tell the relative difference between the individuals, this variable is measured at the ordinal level.

When it comes to the measurement of attitudes among individuals, ordinal variables are almost always used. Questions gauging approval or disapproval of government policies or social behaviors—handgun registration laws, immigration reform, welfare spending, abortion rights, homosexuality, marijuana use, child-rearing practices, and virtually any others that you can think of—are almost always framed by ordinal values.

2.2.3 Interval-Level Variables

Interval variables give the most precise measurements. An **interval-level variable** communicates exact differences between units of analysis. The precision of an interval-level measurement enables you to calculate the precise difference between two data points.

Age measured in years, for example, is an interval-level variable, since each of its values—18 years, 24 years, 77 years, and so on—measures the exact amount of the characteristic. How much difference exists between a subject with 24 years and a subject with 18 years? Exactly 6 years. Because the values of an interval variable are the exact numeric quantities of the measured characteristic, the variable's values do not need to be represented by a separate set of numeric codes. What would be the point? The values themselves tell you all you need to know. If someone were to ask you, "What distance do you drive each day?" your response could be gauged easily by an interval-level value, such as "16 miles." Notice that this value is not simply a number. It is a number that communicates the exact quantity of the characteristic. The researcher would easily determine that your response is different from someone else's answer (such as "15 miles"), that you drive farther each day (because 16 miles is greater than 15 miles), and that the two responses are separated by exactly one unit (1 mile). Ordinal variables do not have equal unit differences, but interval-level measurements do.

Some interval-level variables have **discrete values**. Discrete values usually count the number of times something has occurred, like the number of times a politician has been re-elected or the number of votes cast in the last election. Discrete values are whole numbers. Other interval-level variables have **continuous values**, which means they can have an infinite number of unique values and can be precisely estimated with decimal places. For example, you could calculate your age down to the millisecond and the distance you drive each day down to the millimeter.

It is not difficult to think of interval-level variables in everyday life: the liquid volume of a can of soda, the number of weeks in a semester, the score of a baseball game, or the percentage of one's time devoted to studying. When political researchers are using aggregate-level units of analysis, interval variables are common as well. A student of state politics might measure the percentage of eligible voters who turned out in the gubernatorial election, the number of days before an election that state citizens may register, or the size of the state's education budget. A student of comparative politics might record the number of years that the same regime has been in power in a country or the percentage of the country's budget spent on national defense. A student of interest groups may want to know the membership size, the number of years since the group's founding, or the cost of joining.

Interval-level variables are considered the highest level of measurement because their values do everything that nominal and ordinal values do—they allow the researcher to place units of analysis into different categories, and they permit units to be ranked on the measurement—plus they gauge fine differences between units of analysis.[2]

2.2.4 Which Level of Measurement Is Best?

Generally, political scientists want to measure things as precisely as possible, which means interval-level measurements are ideal. As we'll see, an interval-level variable can easily be transformed into an ordinal- or nominal-level variable if the situation calls for less precision, but it's not a simple matter to make a measurement more precise. However, some things cannot be measured with a great deal of precision. Sometimes, measuring something with a great deal of precision is impractical and is not all that useful. If you were studying the effect of economic status on political behavior, would it really help to have subjects' taxable income calculated to the cent, or would it be sufficient to know what rung of the economic ladder they're on? When political researchers are analyzing individual-level units of analysis—most commonly, individual people—nominal and ordinal variables are much more common than interval variables.

When you do your own political analysis, you'll quickly discover that many concepts and empirical properties can be measured more than one way. For example, marital status is a nominal-level variable, but a researcher interested in the effect of marital status on political behavior could ask survey respondents how many years they have been married. This would be an interval-level measurement of marriage (the value for all unmarried respondents would be 0). Or, to simplify comparisons, the researcher might create an ordinal-level measurement that distinguishes respondents who have never been married, those married less than 10 years, and those married more than 10 years.[3] Similarly, a public opinion scholar could measure an individual's policy preference using a simple dichotomous variable (e.g., favor or oppose), an ordinal-level scale, or an interval-level feeling thermometer that measures an individual's preferences. Depending on the scope and nature of their work, researchers might measure a concept of interest several different ways.

While more precision is generally preferable to less precision, in some situations, less precise measurements are more reliable. For example, many surveys ask respondents about their individual and/or household income. The respondent may be reluctant to supply this personal information, may not know the precise answer, or may have difficulty determining whether gifts, child support, tips, scholarships, and so on are "income" for purposes of answering the question. For these reasons, it may be better to ask respondents to identify their income range on an ordinal scale than to ask them to provide an exact dollar amount.

Some of the most effective measurement strategies in political science are the simplest to understand and apply. One of our favorites is a method used to determine whether a political candidate is a "quality challenger." The concept of a quality challenger is so complicated that it's hard to operationalize it, but Gary Jacobson boiled it down to determining whether the candidate previously held elected office.[4] It's a simple yet effective measurement strategy. Another of our favorite measures is using the front page of the *New York Times* to gauge the salience of an event. If, for example, an oral argument before the U.S. Supreme Court was reported on the front page of the paper, people were paying attention to it.[5] It's a great litmus test of salience because it's easy to understand how it works, it's easy to apply, and the *Times* has been publishing news stories on its front page, with relatively consistent standards, since 1851. This simple measure of salience covers the entire modern political era.

2.3 CENTRAL TENDENCY AND DISPERSION OF VARIABLES

The best understood descriptive statistic is a familiar denizen of everyday life: the average. The world seems defined by averages. When your college or university wants to summarize your entire academic career, what one number does it use? What is the average tuition cost of higher education institutions in your state? When people go on vacation, how many days do they typically stay? What is the most popular month for weddings? What make of automobile do most people drive?

Political research, too, has a passion for identifying what's typical. How much does a congressional candidate commonly spend on a campaign? Do people who describe themselves as Republicans have higher incomes, on average, than Democrats do? What opinion do most people hold on government-subsidized health care? Military spending? Immigration reform?

When it comes to describing variables, averages are indispensable. However, political researchers rarely use the term *average* in the same way it is used in ordinary language. They refer to a variable's **central tendency**—that is, the variable's typical or average value. A variable's central tendency is measured in three ways: the mode, the median, or the mean. The appropriate measure of central tendency depends on the variable's level of measurement.

The most basic measure of central tendency is the **mode**. The mode of a variable is the most common value of the variable, the value that contains the largest number of cases or units of analysis. The mode may be used to describe the central tendency of any variable. For nominal-level variables, however, it is the only measure that may be used.

For describing variables with higher levels of measurement—that is, ordinal or interval—the **median** comes into play. The median is the value of a variable that divides the cases right down the middle—with half the cases having values below and half having values above the median. The central tendency of an ordinal-level variable may be measured by the mode or median.

For interval-level variables, a third measure, the **mean**, also may be used to describe central tendency. The mean comes closest to the everyday use of the term *average*. In fact, a variable's mean *is* its arithmetic average. When we sum all the cases' individual values on a variable and divide by the number of cases, we arrive at the variable's mean value. All these measures of central tendency—the mode, the median, and the mean—are workhorses of description, and they are the main elements in making comparisons and testing hypotheses.

Yet there is more to describing a variable than reporting its measure of central tendency. We can also describe a variable by its **dispersion**, the variation or spread of cases across its values. A variable's dispersion tells us the degree to which observations share the same value or have diverse values. If we're studying individuals, are they alike or different with respect to the characteristic being measured?

A variable's dispersion is sometimes its most interesting and distinctive feature. When we say that opinions on gun control are "polarized," for example, we are describing their variation, the particular way opinions are distributed across the values of the variable—many people support gun control, many people oppose it, and only a few take a middle position. To say that general "consensus" exists among Americans that capitalism is preferable to communism is to denote little variation among people or widespread agreement on one option over another. When scholars of comparative politics discuss the level of economic equality in a country, they are interested in the variation or dispersion of wealth. Is there little variation, with most economic resources being controlled by a few? Or is the distribution more equal, with economic resources dispersed across many or most citizens? Compared with the overworked average—the go-to summary and simplifier—references to a variable's dispersion are uncommon if not rare in everyday life. Variation is underemphasized in social science, too.[6]

In this chapter, we discuss the meaning and appropriate uses of the measures of central tendency—mode, median, and mean. We also explore approaches to describing a variable's dispersion. When we describe one variable at a time, we are working with **univariate statistics**. We are describing the characteristics of one particular variable without considering its relationships with other variables. In a later chapter, we will discuss and analyze relationships between two variables (bivariate statistics), but we can't do that without the vocabulary and tools needed to describe single variables.

2.4 DESCRIBING NOMINAL-LEVEL VARIABLES

A good example of a nominal-level variable used in political science is what Americans consider the "most important problem" facing the country at any given time. Table 2-1, based a March 2023 NPR/PBS/Marist survey of 1,327 adults, shows what Americans consider the most important problem currently facing the country.[7] Table 2-1 is a **frequency distribution**, a tabular summary of a variable's values. Frequency distributions are commonly used in data presentations of all kinds—from survey research and journalistic

TABLE 2-1 ■ Frequency Distribution of Americans' Perception of Most Important Problem

Problem	Frequency	Percentage
Economy	411	31%
Preserving Democracy	265	20%
Health Care	119	9%
Immigration	106	8%
Climate Change	106	8%
Crime	93	7%
Gun Policy	80	6%
Abortion	66	5%
Education	53	4%
Total	1,327	100%

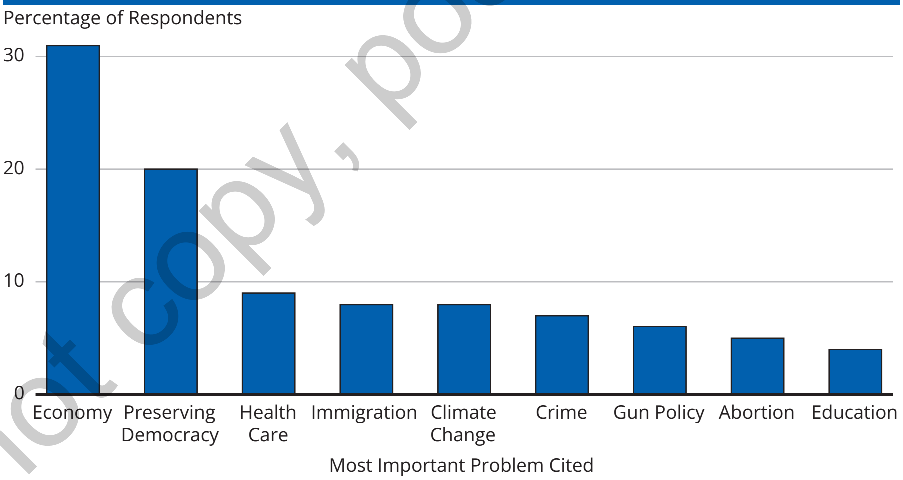
Source: March 2023 NPR/PBS/Marist Poll.

polls to marketing studies and corporate annual reports. The first column of each frequency distribution lists the variable's values. The second column reports the count, or **raw frequency**, of individuals at each of the variable's values. The raw frequencies are totaled at the bottom of the column. This is the **total frequency**. The third column reports the percentage of cases falling into each value of the variable.

Consider the information about central tendency and dispersion of the most important problem variable conveyed by the tabular numbers and graphic display. We can see that the mode is the economy, which is the most important problem identified by 411 people surveyed, or 31% of the sample.[8] Note that the mode is not a percentage or a frequency. The mode is always a value. A good description takes the following form: "Among the [units of analysis], the mode is [modal value], with [percentage of units] having this value." In the example: "Among the 1,327 individuals who identified America's most important problem for this survey, the mode is economy, with 31% of responses having this value."

A picture, to use an old cliché, is worth a thousand words. This adage applies aptly to frequency distributions, which are often presented in the form of a **bar chart**, a graphic display of data. Figure 2-2 shows a bar chart for Americans' perception of the most important problem facing the country (in March 2023). Bar charts are visually pleasing and elegant. The variable's values are labelled along the horizontal axis, and percentages (or, alternatively, raw frequencies) are along the vertical axis. The height of each bar clearly depicts the percentage of respondents citing each problem as most important.[9]

FIGURE 2-2 ■ Bar Chart of Americans' Perception of Most Important Problem



Source: March 2023 NPR/PBS/Marist Poll.

Clearly, the modal value is economy, but would it be accurate to say that the *typical* adult things the economy is the most important problem? Not quite. According to the data, 31% of American adults think the economy is America's most important problem, but this is less than a majority. Preserving democracy is also a frequent choice (20% of respondents).

We organized the values in the table and chart by descending frequency. We show the most commonly cited problem first, the second-most-cited problem next, and so on to help convey what Americans consider the most important problem. But here's the thing about nominal-level variables: There is no inherent order to the values. We could have displayed the values in alphabetical order, by ascending frequency, in the order they appear on the questionnaire. Americans simply have different opinions about their country's most important problem.

Charts help us communicate the most important features of data effectively. They should always be clearly labeled. Avoid using pie charts to display the relative frequency of a variable's values. It's much easier to compare bar heights than radians of pie slices. Academic researchers generally abhor pie charts. We don't use them in this book and recommend avoiding them.

Here is a general rule that applies to any variable at any level of measurement: The greatest amount of dispersion occurs when the cases are equally spread among all values of the variable. As dispersion increases, it becomes increasingly difficult to accurately identify a variable's central tendency, its typical or "average" value. If one-ninth (approximately 11%) of the cases fell into each category of the most important problem variable, this variable would have maximum dispersion. The cases would be evenly spread across all values of the variable. All the bars in the bar chart would have equal heights. Conversely, the lowest amount of dispersion occurs when all the cases are in one value of the variable. If everyone cited the same problem as most important, this variable would have no dispersion at all. The chart would have one bar containing 100% of the cases. Which scenario better describes the most important problem, the equal-percentages-in-each-value scenario or the 100%-in-one-value scenario? Real-world variables rarely fit either scenario perfectly. The most important problem is an example of a variable that has a fairly large amount of dispersion.

To avoid confusion in terminology, we should note that a **proportion** is the raw frequency divided by the total frequency. A **percentage** is a proportion multiplied by 100. Barring rounding error, proportions total to 1.00, and percentages total to 100. Thus, the equation to figure the percentage for each value is as follows:

$$\text{Percentage for each value} = \frac{\text{Raw frequency}}{\text{Total frequency}} \times 100$$

It's important to be able to express frequencies in both percentages and proportions. People tend to understand results expressed in percentages more readily because they're used to percentages. However, when you conduct analysis, you'll often want to work with proportions because it's easier to do math with them. So get used to translating proportions into percentages and percentages into proportions.

2.5 DESCRIBING ORDINAL-LEVEL VARIABLES

Ordinal-level variables are measured with more precision than nominal-level variables, allowing us to make more sophisticated descriptions of their central tendency and variation. Unlike nominal variables, which identify differences among units of analysis, ordinal variables tell us the relative amount of the measured characteristic. Since we can order or rank units of analysis according to the relative amount of the measured characteristic, the order of the rows in a frequency distribution table or bars in a bar chart must be consistent with the relative rank of a variable's values.

Now consider Table 2-2 and Figure 2-3, a frequency distribution and complementary bar chart that display whether Americans think it is easier or harder for people to improve their financial well-being compared to 20 years ago. These data were collected as part of the 2020 American National Election Study. Examine Table 2-2 for a few moments. The frequency and percentage columns have the same meaning as before (see Section 2.4 to review). As with nominal-level variables, we can determine the mode. Here, the modal value is "A great deal harder," which 29.6% of respondents choose.

With ordinal-level variables, we can add a column labeled "cumulative percentage" to a frequency distribution table. The **cumulative percentage** records the percentage of cases at or below any given value of the variable. Thus, 8.1% of respondents think it is now a great deal easier to get ahead now, and 15.4% of respondents think it is a great deal or moderately easier to get ahead now.

TABLE 2-2 ■ Frequency Distribution of Opinions about Ability to Get Ahead Financially

Ability to Get Ahead	Frequency	Percentage	Cumulative Percentage
A great deal easier	589	8.1	8.1
A moderate amount easier	532	7.3	15.4
A little easier	270	3.7	19.1
The same	1,466	20.2	39.3
A little harder	1,090	15.0	54.3
A moderate amount harder	1,167	16.1	70.4
A great deal harder	2,148	29.6	100.0
Total	7,263	100.0	N/A

Source: 2020 American National Election Study

Note: Question: "When it comes to people trying to improve their financial well-being, do you think it is now easier, harder, or the same as it was 20 years ago?"

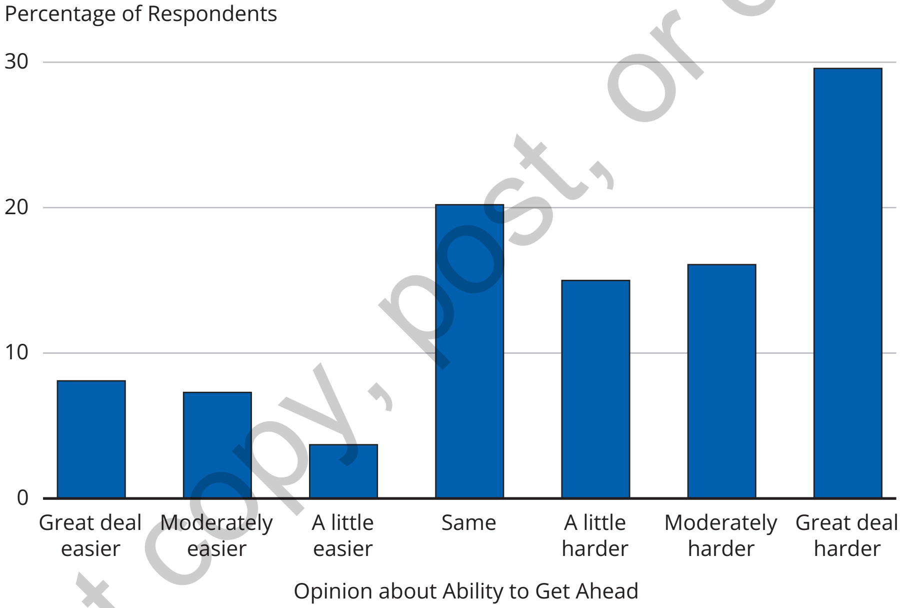
Using the cumulative percentage column, we can locate the median, the value of the variable that (as closely as possible) divides the cases into two equal-sized groups. For ordinal variables—and for interval variables, too—the median value bisects the cases into equal percentages, with 50% of the cases having higher values of the variable and 50% having lower values. The middle-most value of a variable is the median.

What is the median opinion about the ability to get ahead now? This is where the cumulative percentage column of Table 2-2 comes into play. We can see that "A great deal easier" is not the median, since only 8.1% of respondents gave this answer; nor is it "A moderate amount easier," since 15.4% lie at or below this value. To find the median, we work down the cumulative percentage column and stop on the first value that hits 50.0%. The median is within "A little harder" because 54.3% of the cases fall in or below this value.[10]

The median is a specialized member of a larger family of locational measures referred to as percentiles or quantiles. Anyone who has taken a standardized college-entrance exam, such as the SAT, is familiar with this family. A **percentile** reports the percentage of cases in a distribution that lie below it. This information serves to locate the position of an individual value relative to all other values. If a prospective college entrant's SAT score puts him in, say, the 85th percentile on the SAT, that person knows that 85% of all other test-takers had lower scores on the exam (and 15% had higher scores). The median is simply the 50th percentile, the value that divides a distribution in half.

The responses to this question about economic mobility are ordered. Table 2-2 and Figure 2-3 order the variable's value by increasing difficulty to get ahead, starting with a great deal easier and ending with a great deal harder. We could flip the order of responses; we would order them by decreasing difficulty to get ahead, starting with a great deal harder and ending with a great deal easier. The ordering we show is more convenient for analyzing opinions about inequality, while the reverse order is better suited to analyzing opinions about economic mobility. (We discuss this style of variable transformation further in Chapter 3.) The median value is the same either way; the midpoint is within the "A little harder" response. But we can't report other quantiles because the 25% or 75% value depends on whether we order opinions by increasing or decreasing difficulty.

FIGURE 2-3 ■ Bar Chart of Opinions About Ability to Get Ahead Financially



Source: 2020 American National Election Study

How would you describe the distribution of Americans' opinions about the ability to get ahead financially now compared to 20 years ago? The bar chart shows the distribution of responses is heavily skewed to the "harder to get ahead" side. Do the measures of central tendency, the mode and the median, accurately capture its essential characteristics? It's hard to say the variable's central tendency is conveyed by the mode or median because the mode is "a great deal harder," while the median is "a little harder."

Recall the maximum-variation versus no-variation scenarios described earlier. If opinions about economic mobility had maximum dispersion, then equal percentages of respondents would fall into each value. If, by contrast, the variable had no dispersion, then one response category would contain all the cases. By these guidelines, we would conclude that opinions about economic mobility are closer to the maximum-variation pole than to the no-variation pole. Another indication of dispersion, for ordinal variables at least, is provided by comparing the mode and median. If the mode and median are separated by more than one value, then the cases are more spread out than if the mode and median fall in the same value of the variable. By this guideline, too, opinions about economic mobility can be said to have a high degree of dispersion.

Let's describe the observed values of another ordinal-level variable. The same survey that asked Americans their opinions about the ability to get ahead financially asked respondents whether they approve of the president's handling of the economy. Respondents were given four options: approve strongly, approve not strongly, disapprove not strongly, and disapprove strongly.[11] Table 2-3 shows the frequency distribution of responses. Figure 2-4 presents the distribution in a bar chart. How would you describe this variable's central tendency? How about its dispersion?

TABLE 2-3 ■ Frequency Distribution of Approval of President's Handling of Economy

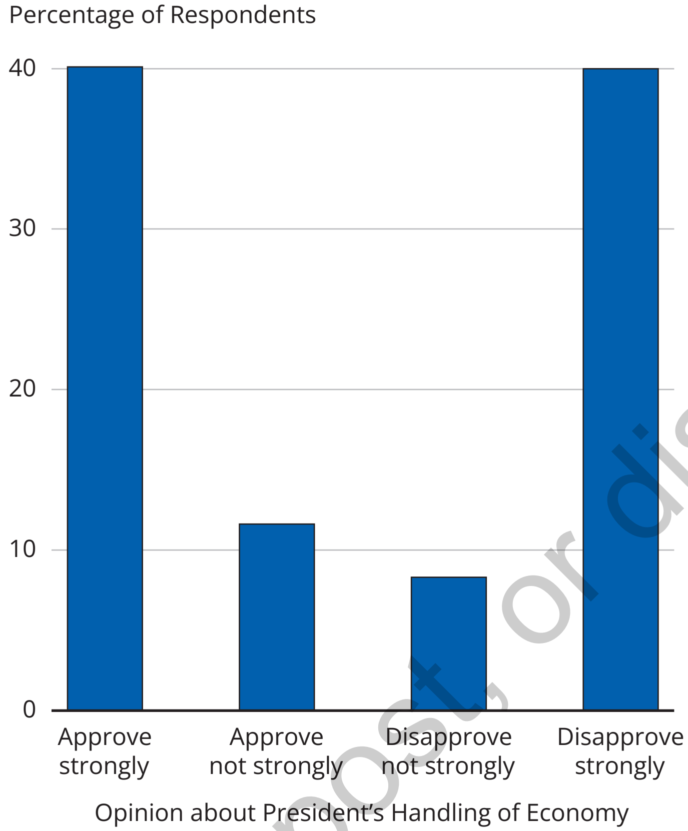
Ideology	Frequency	Percentage	Cumulative Percentage
Approve strongly	3,300.7	40.1	40.1
Approve not strongly	953.5	11.6	51.7
Disapprove not strongly	682.9	8.3	60.0
Disapprove strongly	3,294.9	40.0	100.0
Total	8,232.0	100.0	NA

Source: 2020 American National Election Study

A frequency distribution having two different values that are heavily populated with cases is called a **bimodal distribution**. Americans' opinions about then-President Trump's handling of the economy is a bimodal variable, and a rather interesting one at that. The percentages of the two most extreme values, approve strongly and disapprove strongly, are nearly identical, 40.1% and 40.0%. Also, the two modes are separated by more than one nonmodal category. Because the two modes of this variable are separated by two response categories ("approve not strongly" and "disapprove not strongly"), the distribution is clearly bimodal. We would not want to use a single mode to describe the central tendency of this distribution.

What's the median opinion about the president's handling of the economy? Again, we can move down the cumulative percentage values and stop when it passes 50.0%. Here, the median value is "approve not strongly." Even if we reversed the order to display rows by increasing approval rather than increasing disapproval, the median value would still be "approve not strongly." It is hard to say this is the typical response because only 11.6% of respondents chose it, but it if we think the extreme opinions more or less cancel each other out, the survey suggests modest approval of the president's handling of the economy.

FIGURE 2-4 ■ Bar Chart of Approval of President's Handling of the Economy



Parametric and Non-Parametric Statistics: What's the Difference?

As you proceed in the world of political analysis, you are likely to hear references to parametric and non-parametric statistics. These may sound like advanced electives for statistics majors, but we're using both parametric and non-parametric statistics to describe variables in just the second chapter of this book! Our transition from describing nominal- and ordinal-level variables to describing interval-level variables is a good time to define and distinguish parametric and non-parametric statistics.

Thus far in this chapter, we have analyzed qualitative data measured at the nominal or ordinal level. When we analyze non-numerical data, like Americans' opinions about the nation's most important problem or an ordinal measure of the difficulty of getting ahead financially, we use methods called **non-parametric statistics**. These methods use categories or rank orderings of the data rather than mathematical equations. When a variable conveys non-numeric, qualitative information, we can describe the distribution of its values with a frequency distribution table, we can identify its modal value, and we can depict its distribution with a bar chart. If the variable is ordinal, we can also identify its median value and rank-order the variable's values.

Non-parametric statistics provide robust and flexible tools for data analysis. They work with any type of data, but they are generally less powerful than parametric statistics. Non-parametric statistics work without making strong assumptions about the data and are often used when the data don't follow specific patterns or distributions.

Parametric statistics describe and summarize quantitative and numerical data. These statistics are based on assumptions about the distribution of the variable's values. For example, a researcher might assume that a variable's numeric values follow some well-known statistical distribution, like the bell-curve-shaped normal distribution (discussed in detail in Chapter 8). Statistical distributions, like the normal distributions, are defined by their parameters; normal distributions, for example, are defined by two parameters, mean and standard deviation. So parametric statistics are methods that help us identify the parameters of a statistical distribution that describes a variable's values.

In Section 2.6, we discuss some parametric statistics used to describe interval-level variables. Chances are you've already used some of these tools, even if you didn't think of them as parametric descriptive statistics. When a variable conveys quantitative, numeric information measures at the interval level, we can calculate its mean value using a mathematical equation. We can also apply mathematical formulas to calculate standard deviations and other statistics to describe the distribution of the variable's values. In simple terms, parametric statistics assume specific things about the data and use mathematical formulas to analyze it. They work best when the data follow certain patterns or distributions.

2.6 DESCRIBING INTERVAL-LEVEL VARIABLES

Recall that an interval-level variable gives us precise measurements. Unlike nominal variables, whose values merely represent differences, and ordinal variables, whose values can be ranked, the values of interval variables communicate the exact amount of the characteristic being measured. This means an interval-level measurement can be used to separate cases into groups, rank cases, and calculate differences among cases. What is more, since interval-level variables are the highest form of measurement, each of the "lower" measures of central tendency—the mode and median—also may be used to describe them. When a variable is measured at the interval level, our toolkit for describing the variable's values is extensive.

To illustrate methods used to describe interval-level variables, we will examine and describe poverty rates observed in the 50 states. A state's poverty rate is typically measured by the percentage of people in a state who fall below a specific household income threshold. We're analyzing poverty rate figures rounded to one digit after the decimal place, like 8.4 and 10.7, but measured precisely enough, *no two states have the exact same poverty rate*. This is an important point that affects the tools we can use to describe poverty rates.

2.6.1 Describing Distribution of Values With Tables and Graphs

Table 2-4 reports a frequency distribution for poverty rates observed in the United States. This table looks similar to Tables 2-2 and 2-3; it, too, has columns for frequencies, percentages, and cumulative percentages. But the continuous nature of state poverty rates requires us to adjust the row values. Where Tables 2-2 and 2-3 reported each value of the variables they described on a different row, Table 2-3 describes the frequency of observations within specific intervals. Table 2-4's first row, for example, describes states observed in the 6% to 8% poverty range; there is one state in the 6% to 8% interval. The next row describes states in the 8% to 10% range; there are 11 states with 8% to 10% poverty rates (22% of all states). The intervals used in Table 2-4 are all 2 percentage points wide, which yields a comfortable amount of detail, but you can adjust interval widths to report an interval-level variable's distribution in more or less detail.

TABLE 2-4 ■ Frequency Distribution of Poverty Rates in the United States

Poverty Rate	Frequency	Percentage	Cumulative Percentage
[6–8]	1	2.00	2.00
(8–10]	11	22.00	24.00
(10–12]	17	34.00	58.00
(12–14]	13	26.00	84.00
(14–16]	3	6.00	90.00
(16–18]	2	4.00	94.00
(18–20]	3	6.00	100.00
Total	50	100.00	NA

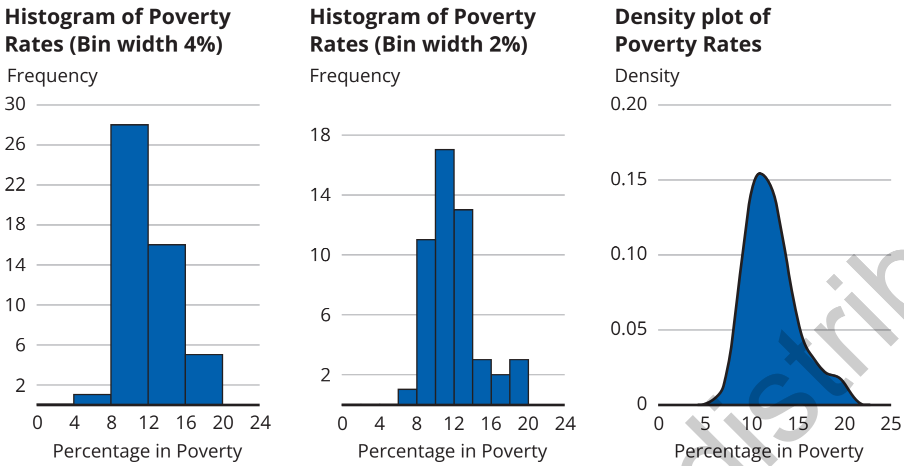
Source: States Dataset (available with *Companions* to Political Analysis)

We've already discussed making bar charts that show the relative frequency of different values of a variable. As mentioned earlier, no two states have the exact same poverty rate; every state has a unique value. A bar chart, then, would show 50 bars with the same height, which is not very informative.

Histograms offer another method of graphing the distribution of an interval-level variable with many unique values. Whereas a bar chart shows the percentages (or frequency) of cases with each value of a variable, a **histogram** shows the percentage or frequency of cases falling into intervals of the variable. The height of the histogram bars represents the number of states with values of the variable in each interval. These intervals, called bins, compress the display, removing choppiness and gaps between the bars. The histogram of state poverty rates in Figure 2-5 depicts the distribution of values reported in Table 2-4.

Density plots (also called kernel density plots) are an alternative to histograms for visualizing the distribution of an interval-level variable. Density plots display a "running average" of observations across the range of observed values. The right panel of Figure 2-4 provides a density plot of poverty rates in the 50 states. You can adjust the level of detail displayed on a density plot by modifying how the density line is fit to the variable's values and how smooth the line should be. Adding to the buffet of options, you can even superimpose a density line on a histogram. Histograms and density plots both allow the researcher to "zoom in" to show greater detail and "zoom out" to reveal general patterns in data. When you create histograms or density plots to describe the distribution of an interval-level variable's values, try to display the right amount of detail.

FIGURE 2-5 ■ Histogram and Density Plot of Poverty Rates in the United States



2.6.2 Measures of Central Tendency

We can describe the central tendency of an interval-level variable, like the poverty rates observed in the 50 states, using the variable's mode, median, and mean. For this specific example, the mode of state poverty rates is not meaningful. A computer program may tell us the mode is a value like 10.1, but this is only due to rounding the rates to one digit. When a variable's values are continuous, it does not have a meaningful mode. Continuous variables are quantities that can take on any value within a certain range, and they can be measured with great precision. This is not to say, however, modes are never used to describe interval-level variables; if an interval-level variable has discrete values, its modal value is important to know. Discrete variables have distinct values and often represent counts. For instance, the number of elections a legislator has won is a discrete, interval-level variable since it can have whole-number values. What is the *median* poverty rate in the United States? As always, a variable's median value is its 50th percentile value; the value that divides observations into equal-sized groups. To find the median, we would list poverty rates in ascending order and find the middle value. We show this in what follows. When there are an even-number of observations, like 50 states, we find the two mid-most numbers and take their average. Here, the mid-most poverty rates are 11.5 and 11.8. The median poverty rate, therefore, is 11.65.

7.3	8.9	9.0	9.0	9.2	9.3	9.3	9.4	9.8	9.9
9.9	10.0	10.1	10.1	10.2	10.4	10.6	10.8	10.9	11.2
11.2	11.3	11.4	11.4	11.5	11.8	11.9	11.9	12.0	12.5
12.6	12.7	12.9	13.0	13.0	13.1	13.3	13.5	13.6	13.6
13.8	13.9	15.2	15.5	16.0	16.2	16.3	18.2	19.0	19.6

What is the *mean* poverty rate among the 50 states? As noted at the beginning of the chapter, the mean is the arithmetic center of an interval-level distribution. The mean is obtained by summing the values of all units of analysis and dividing the sum by the number of units.

You're probably already familiar with calculating means, so let's use this as an opportunity to introduce some notation for variables and statistics that we will use throughout the book. The letter x represents a variable. We can add subscripts to x to identify different instances of x, like $x_1, x_2, x_3 \ldots x_n$. The letter n, whether a subscript or by itself, represents the number of observations. We adorn letters, like x, with accents to signify a calculated or estimated value of that entity; we accent x with a bar, $\bar{x}$, to represent the mean of x. The $\sum$ symbol represents summation. To calculate a mean, we sum together all observed values of x and divide the sum by n.

$$\bar{x} = \frac{\sum x_i}{n}$$

In calculating the mean poverty rate, we add up all observed values ($\sum x_i = 607.2$) and then divide this sum by the number of states ($n = 50$). The result: 12.1. The mean poverty rate among states is 12.1%. When a variable is measured at the interval-level, we can calculate and report its mean value to describe its central tendency. This formal notation for variables and statistics may seem like unnecessary complication at first, especially to define something you already know, but it helps to define important concepts clearly and concisely.

2.6.3 Measures of Dispersion: Range, Standard Deviation, and Variance

How do we describe the dispersion of an interval variable? There are several options, including statistics, like variance and standard deviation, along with quantities like range and interquartile range (IQR).

A rough-and-ready measure of dispersion is provided by the **range**, defined as the maximum actual value minus the minimum actual value. The range of the poverty rates variable, 12.3, is the difference between the variable's maximum value, 19.6, and its minimum value, 7.3.

The **interquartile range** is defined as the range of a variable's values that defines the "middle half" of a distribution—the range between the upper boundary of the lowest quartile (which is the same as the 25th percentile) and the lower boundary of the upper quartile (the 75th percentile).[12] The interquartile range of the poverty rates variable, for example, is 3.35 percentage points, the difference between the 75% value (13.45) and the 25% value (10.1). An interval-level variable's interquartile range can be quite informative, especially when comparing the distributions of different variables. Plus, this measure lends itself productively to graphic display.

For interval-level variables, **standard deviation** (s_x) summarizes the extent to which the cases in an interval-level distribution fall on or close to the mean of the distribution. In gauging variation in interval-level variables, standard deviation is the measure of choice. As its name implies, standard deviation measures the typical amount of deviation of a variable's values from its mean value. Although it is a more precise measure of dispersion than those applied to nominal and ordinal variables, standard deviation is based on the same general principles. If, on the whole, the individual cases do not deviate very much from the variable's mean, then the standard deviation will be a small number. If, by contrast, the individual cases tend to deviate a great deal from the mean—that is, large differences exist between the values of individual cases and the mean of the distribution—then the standard deviation will be a large number.

How does one calculate the standard deviation of an interval-level variable? Let's break down the steps to this calculation.

Step 1. Calculate each value's deviation from the mean. For each observations, $i = 1, 2, 3 \ldots n$, deviation from the mean equals the mean, symbolized as $\bar{x}$ (x bar), subtracted from the observed value, x_i. Because there are 50 states, there will be 50 deviations from the mean. States with a poverty rate below the mean will have negative deviations; states with a poverty rate above the mean will have positive deviations. A state with a poverty rate equal to the mean will have a deviation of 0. Deviations from the mean provide the starting point for calculating the standard deviation.

Step 2. Square each deviation. All measures of variation in interval-level variables, including the standard deviation, are based on the square of the deviations from the mean of the distribution. Squaring each deviation removes minus signs on negative deviations, those states with poverty rates below the mean. Why perform a calculation that eliminates the difference between positive and negative deviations? Because, in the logic of the standard deviation, both positive and negative deviations contribute to the *variation* in poverty rates around the mean.

Step 3. Sum the squared deviations. The summation of the squared deviations, often called the *total sum of squares* (TSS), can be thought of as an overall summary of the variation in a distribution.

$$\text{Sum of squares} = \sum (x_i - \bar{x})^2$$

Notice that if we summed the unsquared deviations, we'd get 0. When calculated on real-world data with many units of analysis, the total sum of squares is always a large and seemingly meaningless number. However, the summation of the squared deviations becomes important in its own right when we discuss correlation and regression analysis (see Chapter 8).

Step 4. Divide the sum of squared deviations by n – 1 to find the variance. The sum of squared deviations divided by $n - 1$ is known by a statistical name, the **variance**. To calculate the variance for a sample, you would divide the sum of the squared deviations by $n - 1$.

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Notice that variance is sensitive to values that lie far away from the mean. That's the beauty of the variance. If a variable's values cluster close to the mean, then the average of the squared deviations will record the closer clustering. As deviations from the mean increase, then the variance increases, too.

Step 5. Take the square root of the variance. The statistic of current concern, standard deviation, is based on variance. Standard deviation is the square root of variance.

These five steps for calculating the standard deviation of an interval-level variable can be summarized in the following standard deviation formula:

$$\text{Standard deviation} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

After going through the standard deviation calculation, you may wonder why we divide the sum of squared deviations by $n - 1$ to find the variance rather than dividing the sum of squared deviations by n (which may seem more intuitive). By dividing the sum of the squared deviations by $n - 1$ instead of n, we correct for the known tendency of the sample variance to underestimate the population variance.[13] The correction is more pronounced for smaller samples (smaller values of n) than for larger samples (larger values of n).

You may also be wondering whether it would be simpler to use the absolute value of deviations from the mean and bypass the steps of squaring deviations and taking the square root of variance. This would give us another, less widely used measure of dispersion, mean absolute deviation, which will be a lower number than the variable's standard deviation.

Table 2-5 offers a detailed look at how one calculates standard deviation. The observations are index $i = 1, 2, 3 \ldots 50$. The x_i column shows the poverty rate observed in each state. Next, the table shows $x_i - \bar{x}$, the deviation from mean observed in each state. As discussed in Section 2.6.1, the mean poverty rate is 12.1. The last column shows the result of squaring each deviation.

TABLE 2-5 ■ Details of Standard Deviation Calculation for State Poverty Rates

Observation	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	15.5	3.4	11.3
2	10.1	-2.0	4.2
3	13.5	1.4	1.8
4	16.2	4.1	16.5
5	11.8	-0.3	0.1
6	9.3	-2.8	8.1
7	10.0	-2.1	4.6
8	11.3	-0.8	0.7
9	12.7	0.6	0.3
10	13.3	1.2	1.3
Rows 11 – 47 omitted for brevity			
48	16	3.9	14.9
49	10.4	-1.7	3
50	10.1	-2.0	4.2
Sum	607.2	0	356.8
Variance = 356.8 / 49 = 7.3 Standard deviation = $\sqrt{7.3}$ = 2.7			

The "sum" reports the summed totals of observed values, deviations, and squared deviations. As discussed, dividing the sum of observed values (607.2) by the number of observed values (50) will yield the variable's mean (12.1). Dividing the sum of squared deviations (356.8) by the number of observed values less one (49) gives us the variable's variance (7.3). The variable's standard deviation is equal to the square root of variance (2.7). Notice that the sum of deviations from the mean value is 0. The positive and negative deviations cancel each other out.

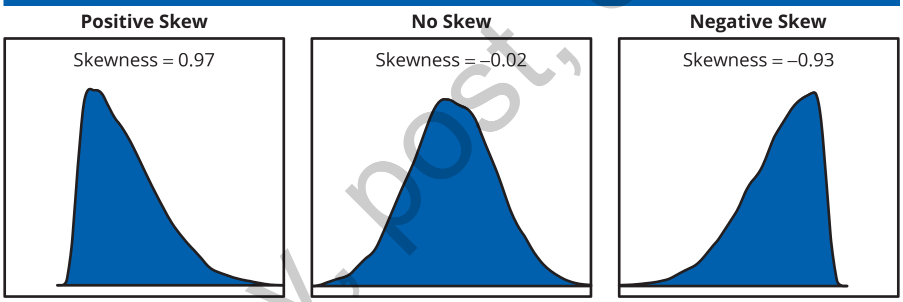
Along with a variable's mean, standard deviation tells you a lot about an interval-level variable. If pressed for time and/or space, report an interval-level variable's mean and standard deviation.

2.6.4 Skewness and Kurtosis

When a variable is measured at the interval-level, we can report some additional statistics to describe the distribution of its values, such as skewness and kurtosis. They are useful statistics. You'll often see them reported in political science research. We'll spare you technical details of calculating these statistics, but we will tell you how we can use them, with the distribution of state poverty rates serving as an example.

Skewness is a measure of symmetry: the more skewed the distribution, the less symmetrical it is. Skewness can be a positive or negative number. Distributions with a longer, or skinnier, right-hand tail have a **positive skew**; those with a skinnier left-hand tail have a **negative skew**. You can see different states of skewness illustrated in Figure 2-6. Generally speaking, when a variable's mean is lower than its median, the distribution has a negative skew; when a variable's mean is higher than the median, the distribution has a positive skew. A distribution that is perfectly symmetrical, one that has no skew, has a skewness equal to zero.

FIGURE 2-6 ■ Distributions With Positive, Negative, and Zero Skewness

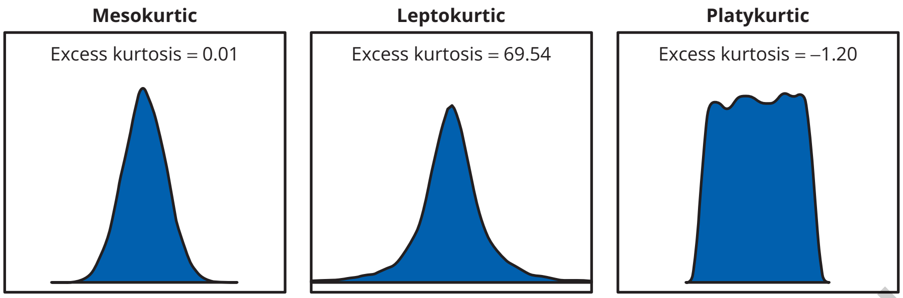


The mean poverty rate is 12.1, and the median is 11.7. The variable's mean is higher than its median. State poverty rates have a positive skew; the variable's skewness is 0.898. Since the mean and median are rarely the same, an interval-level variable will almost always have some skewness. This being the case, should we simply ignore the mean and report only the median of an interval-level variable? How much skewness is too much? Most computer programs provide statistical measures of skewness that can help the analyst in this regard.[14] As a practical matter, however, you have to exercise judgment in deciding how much is too much.

Kurtosis measures the shape of a distribution, specifically how much it deviates from a bell-curve distribution. Kurtosis provides information about the tails and peaks of a distribution and the number of extreme values observed. We won't go into the technical details of kurtosis calculations, but you should know how to read and interpret the statistic when you see it. Like standard deviation, kurtosis is always a positive number. The kurtosis of the poverty rates variable is 3.5. What does this tell us about the distribution of values?

You can see distributions with different types of kurtosis in Figure 2-7. If a variable's kurtosis is greater than 3, it is considered *leptokurtic*; there are more values in the tails

FIGURE 2-7 ■ Distributions With Different Types of Kurtosis



of the distribution, which indicates greater variability and the potential for more rare or extreme events. If a variable's kurtosis is equal to 3, it is considered *mesokurtic*; its distribution closely resembles a bell-shaped curve with a moderate amount of variability. If the variable's kurtosis is less than 3, it is considered *platykurtic*; its distribution has a relatively flat peak and light tails, suggesting less variability and fewer extreme values.

Some computer programs report **excess kurtosis**, which is equal to kurtosis minus 3, which can make it easier to classify the distribution as leptokurtic (positive excess kurtosis), mesokurtic (zero excess kurtosis), or platykurtic (negative excess kurtosis). Applying these definitions to the poverty rates variable, we would classify its distribution as mesokurtic because its kurtosis is greater than 3 (and its excess kurtosis is positive).

2.6.5 Using Box Plots to Compare Dispersions

In this part of the chapter, we will compare the dispersion of some feeling thermometer scores using both descriptive statistics and box plots. You can compare measures of dispersion for feeling thermometer variables because they are measured on the same 0-to-100 scale. Avoid comparing the variances or standard deviations of variables measured with difference metrics.

Take a few moments to familiarize yourself with Table 2-6, which provides summary information for feeling thermometer scores for three government institutions: Congress,

TABLE 2-6 ■ Summary Information for Three Feeling Thermometers

	Congress	U.S. Supreme Court	FBI
Mean	44.8	60.1	62.7
Standard deviation	21.9	22.2	23.7
Mode	50	50	50
N	7,244	7,255	7,140
25th percentile	30	50	50
50th percentile (Median)	50	60	60
75th percentile	60	75	85

Source: 2020 American National Election Study.

Note: Sample weights used for descriptive statistics.

the U.S. Supreme Court, and the FBI. How does the public feel about these three institutions? Notice that these thermometer scales have identical modes (50), but the mean and median feeling thermometer scores are highest for the U.S. Supreme Court (60.1 and 60, respectively) and lowest for Congress (44.8 and 50, respectively). Based on interquartile ranges, feelings about the FBI were more varied than feelings about the Supreme Court or Congress. The IQR for the FBI is 35 points (the difference between the 25% value, 50 points, and the 75% value, 85 points), while the IQR for the Court is 25 points, and the IQR for Congress is 30 points.

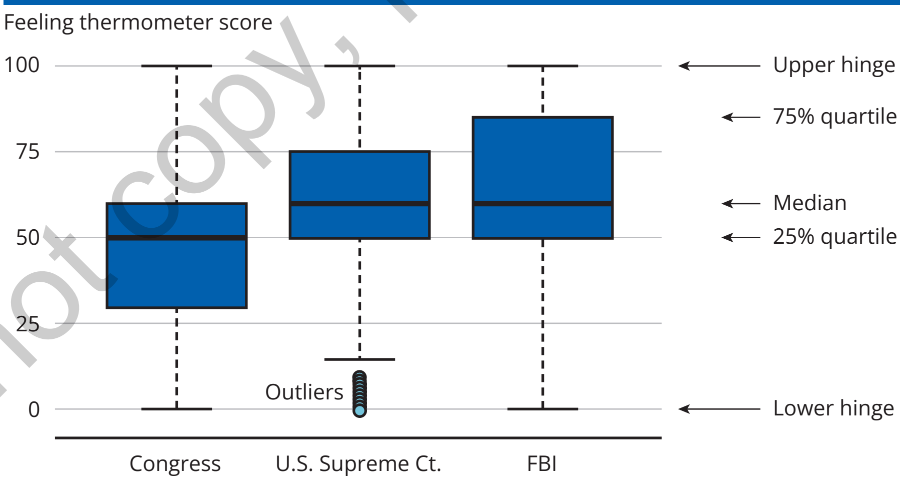
The interquartile range is one element in a box plot, a traditional graphic form that is enjoying something of a renaissance. A **box plot**, which is sometimes called a box-and-whiskers plot, communicates a five-number summary of a variable: minimum value, lower quartile (25th percentile), median, upper quartile (75th percentile), and maximum value.

Figure 2-8 displays box plots for the feeling thermometer scores ANES respondents gave Congress, the U.S. Supreme Court, and the FBI.[15] Box plots of these feeling thermometer scores convey a lot of information about national sentiments about these different institutions.

The shaded boxes convey three values: the lower quartile, the median, and the upper quartile. The boxes identify the IQR for each variable. To display a variable's dispersion, a box plot emphasizes its IQR as opposed to the variable's standard deviation. To display a variable's central tendency, a box plot shows its median rather than its mean.

The lower and upper hinges identify the variable's minimum and maximum values, so long as those values fall within plus or minus 1.5 times the interquartile range of the box. Box plots also identify outliers with markers. Outliers are defined as cases that fall outside the upper and lower hinges.

FIGURE 2-8 ■ Box Plots of Three Feeling Thermometer Variables



Examine the box plot for feeling thermometers scores for the U.S. Supreme Court. The 75% and 25% values for this variable, 75 and 50, define the top and bottom of the shaded box. The horizontal stripe across the box identifies the median, which is 60. The IQR for U.S. Supreme Court ratings is 25 (the difference between the 75% and 25% scores). The upper and lower hinges identify the variable's maximum and minimum values, provided the maximum and minimum are within 37.5 points (1.5 times 25) of the shaded box. This means that the upper hinge extends to 100 (the maximum possible feeling thermometer score), and the lower hinge extends down to 12.5 (37.7 points below the bottom of the box). ANES respondents who give the U.S. Supreme Court scores below 12.5 are outliers marked with hollow circles. You can see that the upper and lower hinges for the feeling thermometers for Congress and the FBI cover the full range of observed values.

Extreme values may have an obvious effect on the mean, but they have little effect on the median. Odd as it may sound, the median is impervious to the amount of variation in a variable. The median reports the value that divides the respondents into equal-sized groups, unfazed by the distribution's skew. For this reason, the median is called a **resistant measure of central tendency**, and you can see why it sometimes gives a more faithful idea of the true center of an interval-level variable.

When a variable is measured at the interval level, our tool kit for describing it is wide open. No single piece of information tells us everything we should know about it. When we describe a variable, we should communicate its most important and interesting features. This often requires us to use our best judgment. In some situations, we can describe a variable better by changing its level of measurement. In the next chapter, we describe some useful methods of transforming variables.

SUMMARY

Variables are perhaps more variable than you had realized before reading this chapter. Table 2-7 provides a thumbnail summary of key differences in variables by level of measurement.

Let's review these points, beginning with the nominal-ordinal-interval distinctions, a persistent source of confusion. The confusion can usually be cleared up by recalling the difference between a variable's name and a variable's values. A variable's name will tell you the characteristic being measured by the variable. But a variable's values will tell you the variable's level of measurement. To figure out a variable's level of measurement, focus on the values and ask yourself this question: Do the values tell me the exact amount of the characteristic being measured? If the answer is yes, then the variable is measured at the interval level. If the answer is no, ask another question: Do the values allow me to say that one unit of analysis has more of the measured characteristic than another unit of analysis? If the answer is yes, then the variable is measured at the ordinal level. If the answer is no, then the variable is measured at the nominal level.

A variable's level of measurement, as we have seen, determines how completely it can be described. We have also seen that describing a variable requires a combination of quantitative knowledge and informed judgment. Table 2-7 offers some general guidelines for interpreting central tendency and dispersion.

For nominal variables, find the mode. Using a bar chart as a visual guide, ask yourself these questions: Is the distribution single peaked with a prominent mode? Is there more than one mode? Visualize what the bar chart would look like if the cases were spread evenly across all values of the variable. What percentage of cases would fall into each value of the variable if it had maximum variation? Compare this mental image to the actual distribution of cases. Would you say that the variable has a large amount of dispersion? A moderate amount? Or are the cases concentrated in the modal value?

TABLE 2-7 ■ Measuring and Describing Variables

	Level of Measurement		
	Nominal	Ordinal	Interval
Precision:			
Separates cases into categories	✓	✓	✓
Allows you to rank cases		✓	✓
Determines the exact amount			✓
Central Tendency:			
Mode	✓	✓	✓
Median		✓	✓
Mean			✓
Dispersion:			
Low if one mode prominent	✓	✓	✓
Low if clustered around median		✓	✓
Low if clustered around mean			✓
High if no clear mode	✓	✓	✓
High if cases spread among values	✓	✓	✓
High if mean and median far apart			✓
Quantified by variance			✓
Quantified by standard deviation			✓
Skew:			
Can be positive or negative		✓	✓
Can be quantified			✓
Graphics:			
Bar chart	✓	✓	✓
Histogram			✓
Density plot			✓
Box plot			✓

For ordinal variables, find the mode and median. Examining the bar chart, mentally construct a few sentences describing the variable. Just as with nominal variables, imagine a maximum dispersion scenario: Does the actual spread of cases across the variable's values approximate maximum variation? With ordinal variables, you also can compare the modal and median values. Are the mode and median the same or very close in value? If so, the central tendency of the variable can be well described by its median. If the mode and median are clearly different values, then it probably would be misleading to make central tendency the focus of description. Instead, describe the variable's dispersion.

For interval variables, find the mode, median, and mean. Because frequency distributions for interval variables tend to be inelegant, a bar chart is essential for getting a clear picture. Consider the three measures of central tendency and examine the shape of the distribution. If the mode, median, and mean fall close to each other on the variable's continuum and the cases tend to cluster around this center of gravity, then use the mean to describe the average value. Just as with nominal and ordinal variables, a diverse spread of cases denotes greater variation. Interval variables also allow evaluations of symmetry. Is the mean a lot higher or lower than the median? If so, then the distribution may be skewed. Describe the source of skewness. Examine the bar chart and decide whether using the mean would convey a distorted picture of the variable. For badly skewed variables, use the median as the best representation of the distribution's center. For interval variables with many values, identify the interquartile range, the range of values between the 25th and 75th percentiles. Use the interquartile range to compare the dispersion of two or more variables that are measured along the same scale.

KEY TERMS

bar chart
bimodal distribution
box plot
central tendency
continuous values
cumulative percentage
density plots
discrete values
dispersion
excess kurtosis
frequency distribution
histogram
interquartile range
interval-level variable
kurtosis
mean
median
mode
negative skew
nominal-level variable
non-parametric statistics
ordinal-level variable
parametric statistics
percentage
percentile
positive skew
proportion
range
raw frequency
resistant measure of central tendency
skewness
standard deviation
total frequency
univariate statistics
variable
variance $\sum(x_i - \bar{x})/(n-1)$

EXERCISES

1. A list of terms follows. For each term, do the following: (i) State whether the term is a variable name or a variable value. (ii) State the level of measurement. *Example:* Support for same-sex marriage. (i) Variable name. (ii) Ordinal.

    - Age
    - Independent-leaner
    - Abortion opinion
    - Taxable income
    - 57 years
    - Millennial
    - Lenient
    - Religious denomination

2. Following are the raw frequencies for two ordinal variables, each of which measures individuals' attitudes toward equality: a measure of egalitarianism (part A) and a measure of tapping level of support for equal pay for men and women (part B). The

egalitarianism variable has three values: low, medium, and high egalitarianism. The equal-pay variable also has three values: people can either favor equal pay for men and women, take a middle position, or oppose equal pay.[16]

For each variable, do the following: (i) Construct a frequency distribution, including frequencies, percentages, and cumulative percentages. (ii) Sketch a bar chart. (iii) Identify the mode. (iv) Identify the median. (v) State whether the variable has high dispersion or low dispersion. (vi) Explain your answer in (v).

A. Raw frequencies for the egalitarianism variable: low egalitarianism, 1,121; medium egalitarianism, 1,133; and high egalitarianism, 1,359.

B. Raw frequencies for the equal pay variable: favor equal pay, 3,187; middle position, 308; and oppose equal pay, 141.

3. A news commentator describes a political candidate, Dewey Cheatum, this way: "Dewey Cheatum is a very polarizing person. People either love him or hate him." Suppose a large number of voting-age adults were asked to rate Dewey Cheatum on an 11-point scale. Respondents could give Cheatum a rating ranging between 0 (they strongly disapprove of him) and 10 (they strongly approve of him).

A. If the commentator's characterization of Dewey Cheatum is correct, what would a bar chart of Cheatum's approval ratings look like? Sketch a bar chart that would fit the commentator's description.

B. Still assuming that the political commentator is correct, which of the following sets of values, Set 1 or Set 2, is more plausible?

☐ Set 1: median, 5; mode, 5.

☐ Set 2: median, 5, two modes (bimodal), 2 and 7.

Explain your choice. Why is the set of numbers you have chosen more plausible than the other set of numbers?

C. Now, suppose that you analyze the actual data and find that the political commentator is, in fact, incorrect. Instead, the following characterization best describes the bar chart of approval ratings: "Dewey Cheatum is a consensus builder, not a polarizer. He generally elicits positive ratings from most people, and there is little variation in these ratings." Sketch a bar chart that would fit this new description. Invent plausible values for the median and mode of this distribution.

4. Following is a horizontal axis that could be used to record the values of an interval-level variable having many values, ranging from low values on the left to high values on the right. Draw and label three horizontal axes just like the one shown here.

Low High

A. Imagine that this variable has a negative skew. What would the distribution of this variable look like if it were negatively skewed? On the first axis you have drawn, sketch a curve depicting a negative skew.

B. Imagine that this variable has a positive skew. On the second axis you have drawn, sketch a curve depicting a positive skew.

C. Imagine that this variable has no skew. On the third axis you have drawn, sketch a curve depicting no skew.

5. Several of this chapter's examples used ANES feeling thermometer scales to illustrate central tendency and dispersion for interval variables. Following is summary information for two more ANES thermometer scales that report ratings for the 2016 vice presidential candidates: Republican Mike Pence and Democrat Tim Kaine.[17] Examine the summaries and think about how the variables are shaped. Consider statements A to E. For each statement, do the following: (i) State whether the statement is true or false. (ii) Explain your answer.

	Mike Pence	Tim Kaine
Mean	48.2	46.0
Standard deviation	29.4	25.9
Mode	50	50
Quartiles:		
25	30	30
50 (Median)	50	50
75	70	60

A. The Tim Kaine feeling thermometer has a negative skew.
B. The mean of the Mike Pence thermometer provides an accurate measure of central tendency.
C. Mike Pence's ratings have a greater amount of variation than Tim Kaine's ratings.
D. A respondent who rated Mike Pence at 58 would have a mean-centered score of about 10.
E. A respondent who rated Tim Kaine at 33 would have a standardized score of 0.5.

6. Seven individuals have been asked to register their opinions on health care reform. Opinions are gauged by a scale that ranges from 0 (the respondent favors a plan based on private medical insurance) to 20 (the respondent favors a government-based plan). The health care opinion scale scores for each individual are as follows: 0, 6, 8, 10, 12, 14, 20.

A. What is the mean health care opinion score for this group?
B. Write down five column headings on a sheet of paper, like the columns shown here: "Respondent," "Health Care Opinion Score," "Deviation From the Mean," "Deviation Squared," and "Standardized Score." The first two columns have been filled in. For this part of the exercise (part B), fill in the two columns labeled "Deviation From the Mean" and "Deviation Squared."

Respondent	Health Care Opinion Score	Deviation from the mean	Deviation Squared	Standardized Score
1	0			
2	6			
3	8			
4	10			
5	12			
6	14			
7	20			

C. Using your calculations in part B, what is the variance? What is the standard deviation?

D. You now have calculated the information you need to fill in the remaining column, "Standardized Score." Go ahead and fill in the remaining column.