

How We Reason

▼ PROLOGUE ▼

Sometimes I wish I were not a political scientist. Unlike some other social sciences, political science also has deep roots in the humanities and in nonquantitative research, and this has led to countless debates in my field concerning epistemology, the branch of philosophy dealing with what is knowledge, and how we should study politics. In a nutshell, the debate is over whether quantitative methods and statistical techniques provide a better picture of reality than more traditional, nonquantitative scholarship.

I have always been a generalist who takes knowledge any way it is available, without dealing with the nature of how we know if what we see actually is what is. Like the late U.S. Supreme Court Justice Potter Stewart, who in 1964 commented that although he couldn't define pornography, he knew it when he saw it, I sometimes wished that the debate would go away and, like psychologists or physicists, we could just "do science," without having to justify what we do to doubting colleagues.

But, on the other hand, maybe there is a role for epistemology. What do we mean when we "do science"? What is science anyway, and how do the social sciences relate to science in general? Before we can "do science," we must know what science is.

INTRODUCTION

This is a textbook on statistics and data analysis for the social sciences. Its techniques apply whenever data that involve counting or measuring have been collected. A standard logical process—the scientific method—underlies the collection and interpretation of data and applies to all the sciences.

The scientific method is the procedure whereby we propose possible relationships among characteristics of phenomena under study and then test to see whether those relationships actually exist. Although the techniques for doing this vary from one discipline to another, the logical sequence remains the same. Therefore, the scientific method provides a good jumping-off point for the chapters to come. This, in other words, is how we reason.

SETTING THE STAGE

All social sciences—indeed, all sciences in general—have many goals. Essentially, we seek to understand some relevant phenomenon to make predictions or provide explanations about that phenomenon and possibly to gain some control over it. To do this, we generally find ourselves concerned with two fundamental kinds of questions—What is? and What ought to be? Questions pertaining to knowing “what is” we call **empirical**, and questions about “what ought to be” we call **normative**. These two types of questions are found side by side in all academic disciplines, with the normative predominating in the humanities, the empirical predominating in the laboratory sciences, and both being components of the social sciences.

Empirical questions Questions that pertain to knowing “what is.”

Normative questions Questions that pertain to “what ought to be.”

Questions about what ought to be form a core essential to understanding our society. We ask such questions as these: What is a good society? What is justice? Is a democratic government an ideal form of political system? Is it moral for opinion leaders to use deceit in dealing with the public? Should we employ the death penalty in the case of premeditated murder? Is racial discrimination morally acceptable? The answers to these questions depend on our values, our priorities and preferences, and our feelings of what is right or just. When we ask, “What are the components of an ideal society?” we are really asking ourselves to specify what we as individuals desire or think contributes to the common good. Normative questions form the

background of social and political philosophy. They extend as far back as human preference can be traced, and some of the greatest minds have sought their answers: Plato, John Locke, John Stuart Mill, Thomas Jefferson, Adam Smith, and others—the giants of normative theory.

The empirical questions pertain to establishment of facts rather than values. We may accept the normative ideal that freedom of the press is a necessity for a free society. The related empirical question would be, Is there freedom of the press in a given nation? and/or How much freedom of the press actually exists there? Suppose we define freedom of the press as the existence of at least one newspaper or TV station not owned or controlled by the government and not subject to censorship by the government. Whatever we may like or dislike about that definition of freedom of the press, once we tentatively agree to use it as our working definition, we are in a position to examine data about each country and, in doing so (perhaps with the aid of experts), to determine whether freedom of the press really exists there. The existence of freedom of the press as we defined it is subject to empirical verification. The question of whether freedom of the press is good and should exist in a democracy is a normative question, a matter of personal preference.

Both normative and empirical considerations enter into any definition of freedom of the press. For instance, we have many examples where countries with wide peacetime press freedoms must impose military censorship in times of war for security purposes. Our values will determine whether we would still categorize that country as having a free press. What about restrictions resulting from libel laws? What about nudity? What kinds of restrictions may we accept and still consider the press to be free? Empirical considerations pertain to our need to be able to clearly measure what we study. How are we going to obtain the information needed to categorize a country's press freedom? If a free press, in our definition, allows "soft-core" but not "hard-core" pornography to be published, how are we to determine where one ends and the other begins?

The normative and the empirical coexist and complement one another in a simple way: Empirical facts tell us how close we are to the normative ideal or how far we still must go to achieve the ideal. Suppose we have normative agreement that poverty should be eliminated. If we know that 10 years ago, 15% of all families were legally defined as living below the poverty level and this year only 12% can be so defined, we might conclude, at least initially, that poverty is declining and we are moving toward our normative goals.

The art of empirical analysis—how we establish facts, how we determine what is, what actually does exist—is the subject matter of this text. Although many techniques are available for studying "what is," an underlying logical process exists to enable us to make use of these analytical techniques.

EXERCISE

For each of the following, indicate which is normative and which is empirical:

1. Abortion should be outlawed.
2. There were 3 abortions per 100 live births in the United States last year.
3. The right to vote is essential to a free society.
4. Only 40% of all registered voters cast votes in the last presidential election.
5. Children should not be physically abused.
6. Last year, the rate of reported child abuse cases rose 3%.
7. A person who is old enough to fight for his or her country and vote in elections should be able to purchase liquor legally.
8. In this state, the minimum age to purchase hard liquor is 21, although the voting age and minimum age for military service is 18.

Obviously, the odd-numbered questions are normative, and their even-numbered counterparts are empirical.

A final observation: That something is normative and therefore exists as a goal rather than an established fact does not make it either more or less valid to scholars than an empirical fact. Both have a place in the way we learn about society. Consider the following:

We hold these truths to be self-evident: That all men are created equal; that they are endowed by their creator with certain unalienable rights; that among these are life, liberty, and the pursuit of happiness. . . .

These are Thomas Jefferson's words from the U.S. Declaration of Independence. "Self-evident" or not, what do we mean by men (persons) being created equal? Equal in size? Intelligence? Strength? Equal before the law? Economically? What about slaves? (Jefferson owned them.) The term *unalienable* means that which cannot be taken away. But your right to life may be taken away by the electric chair, and your right to liberty may be taken away by a prison term (or a required college course).

Does this mean that what is normative is nonsense? Only if we ignore its value in providing guidelines and goals for what ought to be or if we ignore its value in inspiring and motivating us.

By the same token, the following empirical statement may be true factually but hardly inspirational.

A recent survey showed that four out of five consumers prefer Cleeno detergent to the next leading brand.

Unless I were the owner of the company making Cleeno, I would probably be more inspired by the Declaration of Independence than by the detergent statement.

SCIENCE

When we focus our attention on empirical knowledge, we concern ourselves with knowledge based on observation and experimentation. To the extent that we are engaged in discovering and categorizing such empirical knowledge, we are being scientists. **Scientists** are those engaged in collecting and interpreting empirical information. They do so to formulate and test hypotheses. **Hypotheses**, as we use them here, are statements positing possible relationships or associations among the phenomena being studied. These relationships, which will be elaborated on throughout this chapter, suggest that when some attribute or quantity of one phenomenon exists, a specific attribute or quantity of another phenomenon is also likely to occur. Because communications, cultural anthropology, political science, psychology, social work, and sociology, among other disciplines, concern themselves with aspects of societies, they are often termed the **social sciences**. They empirically study social phenomena. Their goals are to formulate and test hypotheses or suppositions about relationships and possible causes and effects among various aspects of a society, a culture, or a political system.

Scientists People who engage in collecting and interpreting empirical information.

Hypotheses Statements positing possible relationships or associations among the phenomena being studied.

Social sciences The empirical study of social phenomena.

Social sciences share with all sciences two common aspects. The first is a commitment to the **scientific method**, a series of logical steps that, if

6 ◀ STATISTICS FOR THE SOCIAL SCIENCES

followed, help minimize any distortion of facts stemming from the researcher's personal values and beliefs. The second is the use of quantitative techniques, measuring and counting, for the gathering and analysis of the factual information that is collected.

Scientific method A series of logical steps that, if followed, help minimize any distortion of facts stemming from the researcher's personal values and beliefs.

The scientific method is really a series of intellectual steps. It is not so much the actual techniques whereby the research is performed as it is the thought process whereby hypotheses are formed, tested, and verified (or not verified). If followed, the scientific method provides a basis for acquiring knowledge that will be eventually accepted by the scientific community. This accepted "truth" would be independent of the values and preferences of the researcher or any other observer. A properly conducted study of attitudes on abortion might find that in a particular group—your class, for instance—55% of the students are pro-choice and 45% are pro-life. This result would be accepted as fact regardless of the researcher's personal preference on the abortion issue.

As we will see, the scientific method involves the formulation of hypotheses, the testing of hypotheses via observation or experimentation, and the ultimate verification (or disconfirmation) of the hypotheses in a manner that will enable other scholars to draw the *same* conclusion as did the investigator. Thus, facts are separated from values.

A major challenge for both the researcher and the consumer of research is to distinguish between what is established fact and what is a moral or ethical value judgment about some aspect of human behavior. This is often not an easy task because both the selection of facts and the interpretation of the selected facts filter through the researcher's own normative screen of values and preferences.

Among scholars, there has long been a philosophical debate over whether a science can truly be value free. (This is a question posed against the claims of those empiricists who desire a value-free science.) Although it is probably true that we can never totally eliminate the effects of value and preference, adherence to the scientific method certainly helps minimize these effects and helps keep those effects from clouding our conclusions. For instance, a medical researcher convinced that cigarette smoke harms nonsmokers may publish a report citing the evidence from studies showing such harmful effects but ignoring those studies showing no damaging

effects. Despite the intervention of personal values in this example, further empirical studies of passive smoking will either verify existing dangerous effects or show that these effects do not exist. Eventually, the issue will be put to rest, despite the occasional encumbrance of personal preference. “Facts are facts.”

To enable us to take a closer look at the scientific method, I have chosen a fundamentally simple example. The example starts with an observation and proceeds to the establishment of theory.

It should be understood that from theory, further observations are generated, leading in time to more elaborate theory. Thus, the process is a cyclical one: observation to theory to observation. Few scientific studies begin without at least some theoretical foundation. In this example, however, we assume no prior theory exists.

Now join me in a literal walk through the scientific method, but bring an umbrella—it might rain.

Example

Suppose each morning I take an hour’s walk. On Sunday, the sun shown during my walk and I admired the blue, cloudless sky. The same was also true on Monday. On Tuesday and again on Wednesday, it was raining and the skies were gray and overcast. On Thursday, it only sprinkled, and the sky, while generally blue, was broken here and there with a dark cloud. Assuming I lack education and prior awareness, I note something obvious to you: Rainfall appears to be associated with the presence of clouds in the sky, whereas sunshine generally means fewer clouds. In any event, whenever it did rain, there were inevitably clouds in the sky. I conclude that rainfall is associated with cloudy conditions. At least that was the case for this 5-day period.

I think to myself that, if for these 5 days, rain is associated with cloudiness, then that same pattern should persist over a longer period. I decide to keep a log, indicating for each day whether I considered it cloudy, partly cloudy, or clear. I also note for each of these days whether or not it rained. At the end of 30 days, I take stock of my results. I note that of the 30 days, 10 were cloudy, 10 were partly cloudy, and 10 were clear. Also, I note that on 15 days, it rained, and on the other 15, it did not. I put together a chart, Table 1.1, to summarize this. We term this chart either a **contingency table**, a **table**, or a **cross-tabulation**. The totals in the margins of my table are called **marginal totals**. The three marginal column totals (10 cloudy, 10 partly cloudy, 10 clear) add up to a **grand total** of 30 days. The two marginal row totals (15 rain, 15 no rain) also add up to the same grand total of 30 days.

Contingency table, table, or cross-tabulation A way of presenting data for purposes of testing hypotheses.

Grand total The total number of cases presented in the table. For instance, in Table 1.1, there are 30 total days being studied.

Table 1.1

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain				15
No Rain				15
Total	10	10	10	30

Now I review my records and tally my results for the 30-day period, noting for each day what the sky conditions were and whether or not it rained (see Table 1.2). I count up my tallies and put the appropriate number in each **cell** (e.g., “cloudy, rain” or “clear, no rain”) in Table 1.3.

Cell Intersection of a particular row and a particular column. For example, in Table 1.3, there are five rainy days with partly cloudy sky conditions.

Table 1.2

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain				15
No Rain				15
Total	10	10	10	30

Table 1.3

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain	10	5	0	15
No Rain	0	5	10	15
Total	10	10	10	30

The results for the 30-day period are consistent with the observations for the initial 5 days: On cloudy days, it always rained; on partly cloudy days, it sometimes rained; and on clear days, it never rained. The presence of rainfall is **associated** with the presence of clouds, and without clouds, it appears that no rain will fall. Similar observations over other 30-day periods of time yield similar results and reinforce my initial conclusions. After a while, I take the conclusion that clouds are associated with rain for granted.

Associated When a case falls into a particular category of one concept, such as rain for the presence of rainfall, falls into a particular category of the other, such as cloudy for sky conditions.

THE SCIENTIFIC METHOD

Aside from my simplistic observational techniques and my simplistic lack of even a basic understanding of weather, in this little example, I have followed the scientific method. This is the thought process that is the underpinning of empirical research. In the first 5 days, I noticed a possible relationship between two **concepts** or ideas: sky conditions and the presence of rainfall. Both of these concepts are known as **variables** because they vary, or change, from one observation to another. Sky conditions vary from cloudy to partly cloudy to clear. Rainfall (as I observed it) varies in the sense that on some days it rains and on other days it does not. (I could also have classified my days more specifically if I had wanted to—for instance, heavy rain, moderate rain, light rain, drizzle, no rain. This time I chose to keep it simple.)

Concepts Ideas.

Variables Concepts that vary, or change, from one observation to another. For instance, some days are rainy; others are not.

The concepts or ideas that we call variables are the phenomena of particular interest in our social science disciplines. They are called variables because they vary in amount or attribute for each individual (or group, or society, or state, or culture—whatever we happen to be observing). Some of the variables (and their categories or amounts) that we may be trying to better understand might include social class (upper, middle, lower), occupational status (white collar, blue collar), political party (Democrat, Republican), status

10 ◀ STATISTICS FOR THE SOCIAL SCIENCES

(ascribed, achieved), government (democratic, authoritarian, totalitarian), or population density (high, medium, low). They parallel the presence of rainfall during my little walks. Other variables may explain the differences in the categories of the variables I wish to better understand: income (high, medium, low—or expressed in actual dollars), education (elementary school, high school, college—or expressed in total years of schooling), religion (Catholic, Protestant, Eastern Orthodox, Jewish, etc.), and type of community (urban, suburban, rural nonfarm, farm). And of course, there may be times when a variable from the first list, such as social class, might be explaining variables in the second list, such as income or education.

The relationship that I noticed on my walk was that certain categories of one variable were associated with specific categories of the other: cloudy sky conditions with the presence of rain, clear sky conditions with no rain. At the end of my initial 5 days of observation, I could have stated what I saw as follows:

There is a relationship between sky conditions and the presence of rain, such that cloudy sky conditions are associated with the presence of rainfall and clear sky conditions are associated with no rainfall.

The above statement we term a **hypothesis**. The hypothesis names the two variables that appear to be related and indicates the nature of that relationship (clouds with rain, no clouds with clear weather). Note that another way of thinking about the relationship in the hypothesis is an “if . . . then” format: If clouds, then there will be rain; if clear, then there will be no rain. Another but wrong hypothesis could have been that clear skies are associated with rain; cloudy skies, with no rain. That would have been an alternative hypothesis, but not one consistent with my initial observations.

Hypothesis A statement that names the variables that appear to be related and indicates the nature of that relationship.

Here are some examples of hypotheses similar to those one would expect to find in the social sciences. These examples make use of some of the concepts presented above:

- ▶ There is a relationship between one’s income and one’s social status, such that the higher one’s social status, the higher will be one’s income, and the lower one’s status, the lower will be one’s income.
- ▶ There is a relationship between the status of one’s occupation and one’s level of education such that individuals with higher status

occupations are more likely to have college educations, and individuals with lower status occupations are more likely to have only grade school educations.

- There is a relationship in the United States between religious preference and partisan identity, such that Democrats draw a greater proportion of support from Catholics and Jews than do Republicans, and Republicans draw a greater proportion of their support from Protestants than do the Democrats.

In each of these cases, despite variations in the wording, the two variables are named, and the relationship between the categories of each variable is specified.

EXERCISE

Try forming some hypotheses that you think may be true using this same basic format.

TESTING HYPOTHESES

Based on a small number of observations during my morning walks (5 days), I noticed a relationship that I then assumed would hold over one or more 30-day periods. In using a small number of observations to assume that the relationship should hold for most or all observations, I was undertaking a process we call **induction**, going from the specific to the general. I induced my hypothesis from five specific observations and then assumed that the hypothesis would apply in all cases.

Induction A process that goes from the specific to the general.

Once the hypothesis was induced, I set out to substantiate the hypothesis by acquiring information over a specific 30-day period. I reasoned that if the hypothesis was true in general, it should be true for the specific 30-day period that I had chosen. Here, I reversed my reasoning and went not from the specific to the general but the other way, from the general to the specific, part of a logical process known as **deduction**. I deduced that if the hypothesis were true all the time, it should be true for the 30-day period I had selected to study. In other areas of research, I might have been able to design an **experiment** to test the hypothesis under laboratory-like

conditions. Here, my “experiment” was to select the 30-day period that I did and keep records on cloudiness and rainfall.

Deduction A process that goes from the general to the specific.

Experiment A test of a hypothesis under laboratory-like conditions.

As the information in Table 1.3 indicates, my hypothesis was verified. If many subsequent studies had produced similar results, so that the relationship between presence of clouds and rainfall was widely accepted by the scholarly community and rarely if ever questioned, then my hypothesis would become a scientific law.

Scientific laws are simply hypotheses with a high probability of being correct. There is never absolute certainty about them, despite our use of the word *law*. Two facts that emphasize this point will be covered in some detail later on. First, because most of our data come from sample surveys or randomized experiments, we are generalizing from a smaller group, the sample, to a larger one, called a population. However, we never can be absolutely sure that what was true for our sample is true for the population as a whole. We only estimate that probability. Second, by the rules of formal logic, we never in sampling actually prove anything. Instead, we demonstrate that all other possible alternatives are unlikely to be true, thus leaving us with only one remaining possibility—the thing that we are proving.

Scientific laws Hypotheses verified so often that they have a high probability of being correct.

Despite these problems, scientific laws differ from hypotheses in general because they are accepted as having a high probability of being correct by the scholarly community actively pursuing research in the field to which that scientific law pertains. The law of gravity would be an example. A scientific law is a law because experts in the appropriate area or discipline have reached that conclusion. It is neither the general public nor scholars in non-related fields who determine what scientific laws have validity. Physicists, not sociologists or theologians, determined the validity of gravity.

But the real world rarely cooperates with the researcher as neatly as indicated in Table 1.3. For example, suppose that after 30 days, I tallied my results and what appears in Table 1.4 emerged.

Table 1.4

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain	5	5	5	15
No Rain	5	5	5	15
Total	10	10	10	30

Table 1.5

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain	8	8	8	24
No Rain	2	2	2	6
Total	10	10	10	30

Here, the hypothesis is not true; it is *disconfirmed* rather than confirmed. There appears to be *no* relationship between sky conditions and the presence of rain. Fifty percent of the cloudy days produced rain (5 out of 10 days), but so did 50% of the partly cloudy days and the same percentage of clear days. Regardless of sky conditions, it rained half of the time. Moreover, knowing a day's sky conditions gives us *no* useful information for predicting rainfall. Compare this to the information in Table 1.3. There, if we know that a given day is cloudy, we can predict rain and be 100% correct (all 10 cloudy days produced rain). If the day is clear, we can perfectly predict no rain (on no clear day did it rain). The partly cloudy days have rain 50% of the time only—this is the only category of sky conditions that does not give us perfect predictability. If on a partly cloudy day we predict rain, we know that we can expect to be correct half of the time. But half of the partly cloudy days produced no rain. We have an equal likelihood of being wrong in predicting rain. By contrast, in Table 1.4, cloud conditions are not useful at all for predicting the weather.

At this juncture, note that Table 1.4 indicates a 50–50 chance of rain, regardless of sky conditions. The 50–50 ratio is the result of the fact that there are an equal number of rainy and no-rain days, 15 days each, in this example. One does not need all equal entries to conclude that no relationship exists. Rather, the cell entries need only be proportionate to the marginal totals. Suppose that out of 30 days studied, it rained 24 days. That would be 80% of all days studied. If within each category of sky conditions

14 ◀ STATISTICS FOR THE SOCIAL SCIENCES

it rains 80% of the time, then we also would have no relationship. Assuming 10 days for each of the three weather conditions, that would amount to 8 rainy and 2 no-rain days in each category of sky conditions. This is illustrated in Table 1.5.

The information found in these tables we call **data**. *Data* is the plural form. One single piece of information should be called a piece of data or a **datum**. Often we forget to differentiate singular from plural, but grammatically, we should say “these data” and so on. In Table 1.3, the data confirm the hypothesis; in Tables 1.4 and 1.5, they do not.

Data All the information we use to verify a hypothesis.

Datum or a piece of data A single piece of information.

From time to time, a relationship may be found that is *not* in the predicted direction, as shown in Table 1.6.

Table 1.6

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain	0	5	10	15
No Rain	10	5	0	15
Total	10	10	10	30

Table 1.7

<i>Presence of Rainfall</i>	<i>Sky Conditions</i>			<i>Total</i>
	<i>Cloudy</i>	<i>Partly Cloudy</i>	<i>Clear</i>	
Rain	8	5	0	13
No Rain	2	5	10	17
Total	10	10	10	30

Here there is a relationship and high predictability, but the relationship does not follow what was logically anticipated by the hypothesis. Clear days produce rain; cloudy days do not. As in the case of Table 1.4, the original hypothesis was not verified, but unlike Table 1.4, there is a relationship

between the variables in Table 1.6, only the relationship is the opposite of the one predicted.

One should note that the relationship found in Table 1.3 is very clear-cut. Rarely do results appear so clear-cut. More likely it is a case where a trend is noticeable, even though there are clear examples of days inconsistent with the hypothesis. Note the illustration in Table 1.7. Only 8 of the 10 cloudy days resulted in rain; 2 days were inconsistent with the anticipated results. Nevertheless, the partly cloudy and clear categories remain unaffected. (The marginal totals for the rows have also changed in this example.) The hypothesis has still been verified, even though the results of the study do not produce perfect predictability for cloudy days. We may conclude that if the clouds appear before the rain (clouds come first in time), then cloudy sky conditions are a **necessary** but not **sufficient** condition for rain. No rain falls without the presence of clouds, but the presence of clouds does not *always* result in rain.

We should be aware of this distinction between necessary and sufficient. A **necessary condition** is a condition that must be present in order for some outcome (in this case, rain) to occur. Its presence, however, does not guarantee that the outcome will occur. By comparison, if a **sufficient condition exists**, the predicted outcome will definitely take place. For example, one could argue that poverty is a cause of communist revolutions. Indeed, the presence of poverty motivated Marx, Lenin, and Mao in their writings and strategies, and there was great poverty in prerevolutionary Russia and China. Yet, many impoverished nations have not undergone Marxist revolutions. Why a revolution in Cuba but not in Haiti? Perhaps poverty is necessary but not sufficient for such a revolution. Then, in addition to poverty, one or more other factors may be needed for a revolution, such as a perception of inequality, unmet rising expectations of an end to poverty, or an organized revolutionary movement. If the presence of poverty alone always led to leftist revolution, then it would be both necessary and sufficient. It is also possible that any of several conditions when accompanying poverty can cause revolution; for example, either poverty plus a perception of inequality or poverty plus a charismatic revolutionary leader is sufficient to bring about revolution. When we study hypotheses containing more than two variables, we take the necessary versus sufficient aspect of relationships into particular consideration.

Necessary condition A condition that must be present in order for some outcome (in this case, rain) to occur.

Sufficient condition A condition in which the predicted outcome will definitely take place.

FROM HYPOTHESES TO THEORIES

Much in the same manner as my rainfall study, I could design studies to test my social science hypotheses. Perhaps I am interested in the possible relationship between religion and political party preference. I could prepare an attitude questionnaire and administer it to a randomly selected group of people. One question would ask each person his or her party preference, and another question would tap religious preference. From the results, we could put together a table similar to those in the rainfall example (see Table 1.8).

Table 1.8

<i>Party Preference</i>	<i>Religious Preference</i>				<i>Total</i>
	<i>Protestant</i>	<i>Catholic</i>	<i>Jewish</i>	<i>Eastern Orthodox</i>	
Republican					
Democrat					
Total					

We would then look for discrepancies in the proportion of each religious group expressing preference first for the Republicans and then for the Democrats.

Let us assume that the findings show what we anticipated—Catholics and Jews have a greater tendency to express preference for the Democratic Party than do Protestants, and Protestants have a greater tendency to identify with the Republican Party than do the two other religious groups. This would motivate me to expand my study. Perhaps I might use as an explanatory variable family origin (for example, the country from which the respondents' forebears came when immigrating to the United States). I might look at the communities where the immigrants settled and the freedom and social mobility available to those immigrants. I might look at the impact of the Great Depression or the civil rights movement on the voting patterns of the families of my respondents. If I can establish a number of interrelated hypotheses, each of which partly explains or accounts for differing partisan identities, I would be developing a **theory** of partisan identity. The theory must also account for as many exceptions to the rule as possible. The better the theory, the fewer exceptions there will be.

Theory A set of interrelated hypotheses that together explain some phenomenon such as why one identifies with a particular political party.

Into my theory would pour dozens of observations and hypotheses, some already known and others undiscovered. For example:

1. African Americans were pro-Republican following the Emancipation Proclamation. As a result, however, of post-Reconstruction Jim Crow legislation, African Americans in the South abandoned the Republican Party.
2. Southern Whites, opposed to Lincoln's policies, identified overwhelmingly with the Democrats until the "Solid South" began to crumble in the 1960s due to the civil rights movement and the perceived liberalism of the Kennedy and Johnson administrations.
3. Irish and Italian immigrants (mainly Catholic) as well as Eastern European Jews immigrated to the United States and settled in large cities, most of which were controlled by the Democrats. Their offspring remained Democrat. By contrast, Catholic and Jewish immigrants who settled in Republican-controlled communities (a smaller number) became Republicans.

Eventually, I would have enough explanatory capability to explain why more Catholics were Democrats than Republicans and why those Catholics who were exceptions identified with the GOP. I would know why Protestants had a higher probability of being Republicans than people from the other religion categories. I would know why African and Jewish Americans were generally Democrats. Finally, I might be able to predict and explain future trends, such as an impending change in party affiliation by one of these groups. In short, I would have a theory of partisan identity.

A similar process could take place in the rainfall example. With the hypothesis tested, I would be inclined to elaborate on it and expand my study. I might decide to replace the variable sky conditions by one such as type of clouds present (e.g., cumulus, cumulonimbus, cirrus, etc.) or humidity level. Perhaps I would also examine temperature or other atmospheric conditions. If I can establish a number of interrelated hypotheses, each of which partly explains or accounts for the presence of rainfall, I would have a theory of rainfall. Going to Table 1.7, I would need to develop hypotheses that would, for instance, explain why those 2 cloudy days produced no rain. What else besides just clouds needs to be present if rain is to

fall? Other hypotheses in my theory would explain such exceptions and more about the nature of rainfall.

TYPES OF RELATIONSHIPS

We worded our original hypothesis this way:

There is a relationship between sky conditions and the presence of rain, such that cloudy sky conditions are associated with the presence of rainfall and clear sky conditions are associated with no rainfall.

We named the two variables and went on to specify the nature of the relationship. When both variables are measured in quantities—as amounts rather than as differing attributes—it is possible to simplify the specification of the relationship in our hypothesis. Both variables must be measuring more or less of an amount. Examples would be net income, either in exact dollars or categorized as high, medium, or low; age, in years or categorized as old, middle aged, or young; or liberalism (high, medium, low). Variables that measure attributes rather than amounts, such as gender, religion, region, or race, require us to word the hypotheses as we have done so far. To illustrate such a simplification of our hypothesis, let us recast the categories of our variables so that both clearly indicate amounts or quantities.

Now in Table 1.9, “Amount of Cloudiness” ranges from most (very cloudy) to least (not cloudy), and “Amount of Rainfall” ranges from most rainfall (heavy) to least rainfall (none). The categories of both variables describe differing amounts, and they are in logical sequence, ordered from largest to smallest amounts. Note that the vast majority of the 30 days cluster in the table along a diagonal line from upper left to lower right, indicating that heavier rains are associated with greater amounts of cloudiness and lighter rains are associated with lesser amounts of cloudiness. Finally, as shown in Figure 1.1, “no rain” is associated with “no cloudiness.”

Table 1.9

<i>Amount of Rainfall</i>	<i>Amount of Cloudiness</i>			<i>Total</i>
	<i>Very Cloudy</i>	<i>Partly Cloudy</i>	<i>Not Cloudy</i>	
Heavy	7	1	0	8
Moderate	2	4	0	6
Light	1	4	0	5
None	0	1	10	11
Total	10	10	10	30

Figure 1.1

<i>Amount of Rainfall</i>	<i>Amount of Cloudiness</i>		
	<i>Very Cloudy</i>	<i>Partly Cloudy</i>	<i>Not Cloudy</i>
Heavy	7	1	0
Moderate	2	4	0
Light	1	4	0
None	0	1	10

We call this diagonal line from upper left to lower right the **main diagonal**. In a table with an equal number of rows and columns, the main diagonal would be a straight line; here it only approximates one. When most cases cluster on or near the main diagonal, indicating that greater amounts of one variable are associated with greater amounts of the other and, conversely, less of one with less of the other, we can describe the nature of the relationships with the expression **positively related**. A **positive relationship** is one where greater is associated with greater; less with less.

Main diagonal Diagonal line from upper left to lower right.

Positive relationship A relationship in which greater is associated with greater; less with less.

Our hypothesis may now be reworded in simplified form:

The amount of cloudiness and the amount of rainfall are positively related.

Please note that in each instance, the term *positively* pertains to the nature of the actual relationship (high with high, low with low). It is *not* used as a description of how certain we are that a relationship exists. We are not saying that we are “positive” that the two variables are associated. We are not saying that cloudiness and rainfall are “absolutely, positively” related. Rather, we are hypothesizing that the nature of that association is that the more cloudiness there is, the more rain there will be.

The opposite of a positive relationship occurs when the variables are related in a way that the more of one variable there is, the less of the other there will be. Suppose more clouds mean less rain and fewer clouds mean more rain, as shown in Figure 1.2.

Figure 1.2

<i>Amount of Rainfall</i>	<i>Amount of Cloudiness</i>		
	<i>Very Cloudy</i>	<i>Partly Cloudy</i>	<i>Not Cloudy</i>
Heavy	0	1	10
Moderate	1	4	0
Light	2	4	0
None	7	1	0

In Figure 1.2, the clustering is on a diagonal line going from the upper right-hand side of the table to the lower left-hand side. We refer to this as the **off diagonal**, and when most cases cluster about the off diagonal, we say that the variables are **inversely related**. The term *negatively related* is sometimes used, but the term *inversely related* is preferred.

Off diagonal Clustering on a diagonal line that goes from the upper right-hand side of the table to the lower left-hand side.

Inversely related A condition in which most cases cluster about the off diagonal. A high score on one variable is associated with a low score on the other.

If we were initially inclined to believe clouds were associated with a lack of rain and clear days were associated with rainfall, our hypothesis could have been worded as follows:

The amount of cloudiness and the amount of rainfall are inversely related.

Warning: The nature of a relationship, positive versus inverse, is taken from the logic of the hypothesis and corresponds to the stated diagonals in a table only if the table is set up so that the main diagonal represents “more with more” and “less with less.”

Suppose Figure 1.1 had been recast as in Figure 1.3.

Figure 1.3

<i>Amount of Rainfall</i>	<i>Amount of Cloudiness</i>		
	<i>Very Cloudy</i>	<i>Partly Cloudy</i>	<i>Very Cloudy</i>
Heavy	0	1	7
Moderate	0	4	2
Light	0	4	1
None	10	1	0

Although the categories of rainfall amount remain as before, the other variable, amount of cloudiness, was entered in the sequence opposite the way it was done in Figure 1.1. Now the clustering is on the off diagonal, but the relationship is *positive*, not inverse. Heavy rainfall is still associated with very cloudy sky conditions; low rainfall, with clear days. While the general convention is to set up tables so that clustering in the main diagonal indicates a positive relationship, there are exceptions to every convention. One must be on guard for such exceptions.

Finally, to reiterate a point made earlier, the positive versus inverse relationship terminology cannot be used unless both variables have categories representing *amounts* of the variables! Suppose we have a hypothesis that states that an individual's gender is related to his or her hair color, such that women are more likely to be blondes than men, and men are more likely to have dark hair than women. A study of 100 people yields the data shown in Table 1.10.

Table 1.10

<i>Gender</i>	<i>Hair Color</i>		<i>Total</i>
	<i>Brown or Black</i>	<i>Blonde</i>	
Male	35	15	50
Female	15	35	50
Total	50	50	100

Since 70 of the 100 people appear to be consistent with the hypothesis (men with dark hair; women with blonde hair), the hypothesis is verified.

22 ◀ STATISTICS FOR THE SOCIAL SCIENCES

Yet, we could not say that hair color and gender are “positively associated.” There is no quantification in either variable in the sense of the categories implying more or less of the variable. Male and female are two types of gender; neither category possesses more or less gender. The same applies to hair color. Blonde is perceived by most people as a different color than dark hair, but a blonde has neither more nor less *an amount* of hair color than a dark-haired person. (Ignore the fact that physicists do view colors in amounts—the frequency of light from one end of the spectrum to the other.) Thus, the term *positive* or *inverse* would be unclear in characterizing this relationship and should not be used.

Let us examine how some social science hypotheses might be worded in this new manner:

- ▶ Income and social alienation are inversely related.
- ▶ Occupational status and educational level are positively related.
- ▶ Support for the existing political system and expectations of impending improvement in one’s living standards are positively related.
- ▶ Time spent viewing television and time spent reading print media are inversely related.
- ▶ Income and support for organized labor are inversely related.

We could not word the hypothesis relating religion to partisan identity in this manner, however. Such a hypothesis relates attributes—Republican, Democrat—to other attributes—Protestant, Catholic, and so on. These attributes are not quantitative in nature. They do not represent amounts of the variables.

ASSOCIATION AND CAUSATION

When we do research, we are ultimately seeking to infer the **cause** of the phenomenon we study. What variables when changing in value cause the variable we are studying to change? What factors account for the amount of rainfall? If levels of cloudiness are associated with levels of rainfall, and we assume for a moment that the actions of other variables such as temperature and humidity play no role in the direct relationship between cloudiness and rainfall, we are tempted to suggest a *causal relationship* between the two. We might assume that changes in levels of cloudiness bring about changes in the amount of rainfall. In effect, we are saying that it is the clouds that cause the rain. Given what is known about meteorology and climate, it is a logical assumption that clouds cause rain. If we knew nothing

else about weather, however, we might just as likely conclude that it is the rainfall that causes the cloudiness level. Which of the two directions of causation we choose will often depend on two things: the logic of the situation and the temporal sequence of the variables, or which variable came first in time.

Cause When one phenomenon being studied brings about the other.

As a point of departure, it should be borne in mind that, in general, it is illogical to talk about a pattern of causation between two variables unless it can be first demonstrated that there is association between the variables. Without association, it is meaningless to consider causation. As demonstrated in Table 1.4, the same proportion of rainy days exists under cloudy conditions as under partly cloudy and as under clear conditions. Knowing sky conditions does *not* improve our ability to predict or explain rainfall. Conversely, knowing whether this is a rainy day or not does not help us predict or explain cloud level. If variation in one variable does not relate to variation in the other variable, there is no evidence to suggest that either variable causes, or brings about a change in, the other variable. Thus, with the exception of a few instances presented in subsequent chapters, *association is a necessary condition for causation*.

Once association has been established, it may be possible to infer a causal direction between the two variables. A simple tool for doing this, particularly when we want to explain and not just predict the change in a variable, is **temporal sequence**. If one variable changes earlier in time than does the other variable, we might assume that the former may cause the latter. In the case of the rainfall problem, we may observe that the buildup in the cloud level takes place prior to the rainfall. In fact, once the rains stop, the cloud level often dissipates. Since the cloudiness precedes the rainfall, we may assume that clouds cause the rain. Since the rain follows the cloudiness in time, it would make no sense to suggest that the rainfall causes the cloudiness. The cause precedes the effect temporally.

Temporal sequence When one phenomenon being studied occurs earlier in time than the other.

Unfortunately, we are not always able to ascertain the time ordering of the variables. The gender versus hair color problem presented in Table 1.10

is an example of this dilemma. We receive through inheritance of genetic factors both our gender and hair color predispositions before we are born. From this perspective, it would be just as illogical to assume that gender causes hair color as it would be to assume that hair color causes gender. Even if there is association found between the two variables, there is no evidence that one either preceded the other in time or could logically have caused the other.

Given this fact, it would seem reasonable not to concern ourselves with the causation question at all unless we had evidence for making a causal inference. Unfortunately, many of the statistical techniques used in data analysis require that we designate, *in advance* of calculating the statistic, which variable is doing the causing and which variable is being caused. Because of this problem, there will be times when we may have to arbitrarily select one variable to be doing the causing or explaining and, by default, assume that the other variable's changes are being "caused" by changes in the former. In the case of the gender versus hair color problem, if we are primarily interested in studying hair color, we are likely to assume that for our purposes, gender "causes" hair color. If our goal were to account for gender differences, we would have to treat differences in one's hair color as leading to differences in one's gender.

Likewise, in American politics, there has been a relationship between religion and partisan identity. Although it has been dissipating in recent years, Protestants have had a slightly higher affinity for the Republicans, whereas Catholics and Jews have been more likely to vote Democrat. What is the variable doing the causing? Probably religious identity comes first in time, although not by much, so it would be logical to consider religion the cause and partisanship the effect. This would certainly make sense to a political scientist who would be trying to account for party identification. However, if one's field is the sociology of religion, and religious identity is the subject of inquiry, it would be perfectly logical to treat partisan identity as causing religious identity.

When we must differentiate the variable presumed to do the causing from the variable being caused—whether the selection is based on logic or based on an arbitrary decision—we usually call the variable being described, caused, or explained the **dependent variable**, and the variable doing the causing or explaining is the **independent variable**.

Dependent variable The variable that is being caused or explained.

Independent variable The variable that is doing the causing or explaining.

The dependent variable is the “causee”; the independent variable is the “causer.” Changes in the dependent variable *depend* on changes in the independent variable but not necessarily the other way around, just as increases in rainfall depend on increases in cloud level, but rainfall does not cause the clouds to form.

If we hypothesize that social inequality leads to revolution, then the occurrence of a revolution depends on prior social inequality. Revolution is the dependent variable (being caused), and social inequality is the independent variable (doing the causing). If we believe that air pollution (occurring first in time) causes certain forms of cancer, then level of air pollution is the independent variable, and the cancer rate is the dependent variable. Implicitly, we assume in the latter instance that cancer levels do not cause air pollution.

Some people take issue with the use of the terms *dependent* and *independent variable* in cases where no logical causal ordering can be inferred. They prefer the term **criterion variable** as a substitute for *dependent variable* and **predictor variable** as a replacement for *independent variable*. With this alternate terminology, there is no connotation of causality or implication of change in one variable being dependent on change in the other. Nevertheless, while one should be aware of this alternative to the terms *dependent* and *independent variable*, this text will continue to use the more traditional dependent/independent variable terminology.

Criterion variable A substitute for the term dependent variable.

Predictor variable A substitute for the term independent variable.

THE UNIT OF ANALYSIS

There remains one other item of discussion in this review of the scientific method—the unit of analysis. The **unit of analysis** is what we actually measure or study to test our hypothesis. It is not the variable being studied but rather the entity being studied—the person, place, or thing from which a measurement is obtained. In the rainfall problem, days were the units of analysis. For each of the 30 days, we took two “measurements,” the presence (or amount) of rainfall and the presence (or amount) of clouds. In the hair color/gender problem, individual people were the units of analysis. For each person, we determined two things: that individual’s gender and that individual’s hair color. In the problem asking whether social inequality led to revolution, we would have to design a study in which we collected

information from a number of countries. Thus, country or nation-state would be the unit of analysis. For each country in our study, we would then seek to ascertain its people's level of inequality and also whether revolution had taken place in that country during some specified time interval. That would likely be the way a comparative political scientist would handle the problem. By contrast, a social psychologist might hypothesize that for an individual, his or her self-perception of being the victim of social inequality would influence his or her tendency to be supportive of revolution. Here, individuals, not countries, would be studied, so individuals would be the units of analysis. An urbanist might assume that where murder rates are high, so too would robbery rates be high. Data might be collected from a number of cities, finding each one's murder and robbery rates for a given year. Cities would be the units of analysis.

Unit of analysis What we actually measure or study to test our hypothesis from whom or from what the measurement is made.

Since so many social and behavioral scientists study aspects of human attitudes and behavior, we often encounter an individual of one kind or another being studied as the unit of analysis. Thus, many examples in this book also use the individual as the unit of analysis. It must not be assumed that this is always the case, however. Depending on availability of data, the units of analysis could be individuals; business firms or social organizations; counties, states, or provinces; or nations or international alliances. Care must be taken to avoid assuming that what holds true for one unit of analysis holds true for others. Conclusions about the behavior of business firms do not necessarily carry over to individuals or counties or other units of analysis.

Example

Suppose we were doing a study of some state's criminal justice system. During the process of collecting data, we notice that in the case of public defenders, those people employed in urban areas seem to have higher caseloads than those employed in rural areas. Assuming that we wish to understand *caseload* (amount of cases per public defender) and that a community's *population size* (urban, implying large population, and rural, implying small) may account for the variations in size of public defender caseload, then caseload is our dependent variable, and population size is our independent variable.

The hypothesis we induce from our observations is as follows:

There is a relationship between the size of a public defender's caseload and the size of the community where that public defender is employed, such that public defenders in urban communities have higher caseloads than their colleagues in rural communities.

Noting that both caseload and population size are quantifiable variables, we may simplify our hypothesis as follows:

The size of a public defender's caseload and the size of the community employing that individual are positively related.

To test our hypothesis, suppose we have access to data for each county in that state or province showing the county's average public defender caseload and also that county's population. County is our unit of analysis.

To keep our example *very* simple, assume that we establish a cutoff point in terms of caseload and another cutoff point in terms of population size, such that each county is classified as either high or low in terms of public defender caseload and urban or rural in terms of population size.

If the hypothesis we induced is assumed to be true universally, it is assumed to be true—we deduce—for this particular province or state. We categorize each county in terms of caseload (high versus low) and population (urban versus rural).

Note that our hypothesis is empirical; it can be tested from the data at hand. Whether or not the hypothesis is true is kept apart—as much as we can—from our own normative judgments. The facts will hold, regardless of our own normative opinions and beliefs about what *should* be the case. These normative beliefs could be any of a number of possible attitudes:

Public defenders *ought* to have equal caseloads, regardless of population density, since it is *unfair* for urban case workers to have greater workloads than their country cousins.

Public defenders in cities *ought* to have higher caseloads than their rural counterparts since cities have higher concentrations of poor and disadvantaged, and thus more crime. Therefore, urban personnel *should* be working harder.

Public defenders in rural areas *ought* to have caseloads as high as their more assertive colleagues in the cities.

Public defenders in cities *ought* to have higher caseloads than rural public defenders since the former were foolish enough to select jobs in the cities.

Regardless of what we think ought to be, the empirical study will tell us what actually is the case.

Assume that we are studying 40 counties, of which half are classified as urban and half are rural. A tabulation of our results might look like this:

Table 1.11

	<i>Size of County</i>		
	<i>Caseload</i>	<i>Urban</i>	<i>Rural</i>
High		17	5
Low		3	15
Total		20	20

We note that out of 40 counties, 32 are consistent with the hypothesis (high load/urban or low load/rural). Only 8 counties are classified in a manner inconsistent with the hypothesis. The evidence suggests that our hypothesis is confirmed.

But what about the 8 inconsistent counties? Why did 3 urban counties have low caseloads and 5 rural counties have high caseloads? To move toward a *theory* of caseloads, we need to test the impact of other independent variables on caseload. Which ones might they be? Perhaps the 3 urban counties with low caseloads are prosperous counties with a high tax base. We could study the impact of median income or tax revenue on caseload. Perhaps the political party controlling the county's government has an impact. We could make the governing party our independent variable. Perhaps the nature of the crimes committed in the county can explain some of the inconsistencies. Maybe the 5 rural counties with high caseloads have a large rate of misdemeanors—easily disposed of by the judicial system. Perhaps the other 15 rural counties have a larger occurrence of felonies. The cases may be fewer, but their adjudication might be more complex and time-consuming. What other independent variables might you want to test to build our theory?

Over time, if we repeat these studies in other places and keep getting similar results, confidence will grow as to the validity of our hypothesis. We have, to the extent that science allows, verified our hypothesis empirically and moved along the road of theory building.

CONCLUSION

In this chapter, we have discussed the scientific method. Scientific reasoning is by no means the only way to understand the world. We could view the world through more traditional ways, such as through theology, a political ideology, facts or myths generated by our cultural environment, or simply what those in authority tell us. All of these, however, require faith in the sources telling us about the world and faith in those who interpret those sources for us. Scientific reasoning also requires faith, but it is a faith in ourselves and our colleagues. This is a faith based not on outside authority but on our own ability to collect and interpret data and our ability to scrutinize the research of others and to be able to reach the same conclusions they reached.

The scientific method provides us with logical steps for formulating and testing hypotheses. This thought process parallels the process used in all scientific research and remains stable. What does not remain so stable are the techniques of observation and experimentation used to verify hypotheses, research techniques, and the data analysis techniques used in reaching conclusions.

Research techniques vary with the field of study. In many of the social sciences, we use some observation and experimentation techniques, but we also depend quite a bit on survey research through interviews and questionnaires. Other social sciences such as psychology may use the same techniques but put more emphasis on experimentation.

Although this chapter focuses on the scientific method, the chapters that follow concentrate on the techniques of quantitative analysis—not so much on the research design but on the techniques of measuring and counting for the purposes of analyzing data and showing how the numbers come to tell us what the facts are. All of these topics are part of the field of **statistics**, the study of how we describe and make inferences from data. Our aim will be to learn how to make the numbers make sense.

Statistics The study of how we describe and make inferences from data.

The techniques employed in analyzing data depend in part on the type of research design used and in part on other factors that we will encounter in Chapter 2, “Levels of Measurement and Forms of Data.” Before going on, try working the following exercises.

EXERCISES

Exercise 1.1

1. Develop a hypothesis appropriate to your major field of interest. Word the hypothesis using one of the formats presented in this chapter.
2. Identify the dependent and independent variables. What was the reasoning behind your decision as to which was which?
3. Identify logical categories for each variable. Are you measuring amounts or attributes?
4. What is your unit of analysis that you will need to study, and how will you collect the data?
5. Put together a table similar to the one in the caseload/population example (see page *000*). Simulate the numbers in the table to resemble the results you would expect to get if your hypothesis is verified.

Exercise 1.2

For the three hypotheses on pages *000–000*, repeat Steps 2, 3, and 4 as in Exercise 1.1.

Exercise 1.3

Formulate hypotheses useful for explaining or for measuring progress toward attaining the following normative goals. For each hypothesis, state the dependent and independent variables, their categories (or how you will measure amounts of each variable), and the units of analysis.

1. Women and men should receive equal salaries for similar occupations.
2. Housing restrictions on minority communities should be eliminated.
3. Students studying foreign languages should study the languages of the major linguistic minority groups in their country (e.g., French in Canada or Spanish in the United States).
4. Campaign contributions by interest groups or political action committees should be limited by law.
5. Chemical and biological weapons should be eliminated.
6. The death penalty should be applied in a timely manner.
7. All who are mentally ill should receive treatment.
8. Use of illegal drugs should be eliminated.
9. All citizens should receive a college education.
10. All gun control laws should be repealed.

Exercise 1.4

Each of the following hypotheses has a flaw in either its format or its logic. Identify the flaw and correct the hypothesis.

1. There is a relationship between women and math anxiety such that women have math anxiety.
2. Are age and need for social services positively related?
3. There is a relationship between birth weight and smoking such that mothers who smoke have lower birth weight.
4. There is a relationship between British political parties and support for national health insurance such that Labour, Conservative, and Liberal Democrats support national health insurance.
5. Religion and support for church tax exemption are positively related.
6. Liberals run for public office.
7. Communication graduates earn less than public administration graduates and thus drive cheaper cars.
8. "Right-brained" people are more likely to vote for conservative candidates.

Exercise 1.5

Identify the appropriate unit of analysis for each of the following. Be as specific as possible.

1. Levels of censorship are greater among countries at war than those at peace.
2. Urban areas have higher juvenile delinquency rates than do rural ones.
3. Per capita income is higher in English counties than in the rest of Britain.
4. Two thirds of the kindergarten students in Mrs. Smith's class at Apple Valley Elementary School were absent last February 7, due to the flu.
5. Managers are more likely to contribute to charities than are technicians.
6. First-degree murder rates tend to be higher in southwestern U.S. states than in southeastern ones.
7. Of all NHL teams that year, Detroit averaged the most goals per game.
8. Elvis Presley sold more albums than Jerry Lee Lewis, Bo Diddley, Chuck Berry, or any other musicians of that era.
9. Tokyo and Mexico City are the largest and second largest world metropolitan areas, respectively.
10. The United States had more weapons of mass destruction than did Iraq.

▼ **KEY CONCEPTS** ▼

measurement	frequency distributions	ratio level of measurement
levels of measurement/ scales	dichotomies	absolute zero
qualitative	<i>n</i> -category variables	demographic variables
(categorical) data versus quantitative data	ordinal level of measurement	individual interval data/raw scores
nominal level of measurement	individual ordinal data	ungrouped frequency distribution
attributes versus quantities	grouped ordinal data	grouped interval data
individual nominal data	Likert scale	class interval
grouped nominal data	frequencies scores	closed-ended versus open-ended class intervals
	interval level of measurement	
