where $\sum_{i=1}^{n} y_i$ is a convenient way of writing $y_1 + y_2 + y_3 + \ldots + y_n$. Another statistic is the *sample maximum*, labeled $y_{\max}$:

$$y_{\max} = \text{Maximum}(y_1, y_2, \ldots, y_n) \tag{2.1}$$

The sample mean of the four incomes from the example in section 2.4 (\$9,000, \$22,000, \$21,000, and \$54,292) is \$26,573. The sample maximum is \$54,292. We can summarize the original data containing four numbers with these two numbers representing the sample mean and maximum. We can also calculate other sample characteristics, such as the minimum, median, mode, or variance.

Each summary in this model reduces all the data (four numbers in this simple example, or our knowledge of some aspect of European history in the other) to a single number. Communicating with summaries is often easier and more meaningful to a reader than using all the original data. Of course, if we had only four numbers in a data set, then it would make little sense to use five different summaries; presenting the four original numbers would be simpler. Interpreting a statistic is generally easier than understanding the entire data set, but we necessarily lose information by describing a large set of numbers with only a few.

What rules govern the summary of historical detail? The first rule is that *summaries should focus on the outcomes that we wish to describe or explain.* If we were interested in the growth of the average international organization, we would not be wise to focus on the United Nations; but if we were concerned about the size distribution of international organizations, from big to small, the United Nations would surely be one of the units on which we ought to concentrate. The United Nations is not a representative organization, but it is an important one. In statistical terms, to investigate the typical international organization, we would examine mean values (of budgets, tasks, memberships, etc.), but to understand the range of activity, we would want to examine the variance. A second, equally obvious precept is that *a summary must simplify the information at our disposal.* In quantitative terms, this rule means that we should always use fewer summary statistics than units in the original data, otherwise, we could as easily present all the original data without any summary at all.[6] Our summary should also be sufficiently simple that it can be understood by our audience. No phenomenon can be summarized perfectly, so standards of adequacy must depend on our purposes and on the audience. For ex-

[6] This point is closely related to the concept of indeterminant research designs, which we discuss in section 4.1.

ample, a scientific paper on wars and alliances might include data involving 10,000 observations. In such a paper, summaries of the data using fifty numbers might be justified; however, even for an expert, fifty separate indicators might be incomprehensible without some further summary. For a lecture on the subject to an undergraduate class, three charts might be superior.

2.6 Descriptive Inference

Descriptive inference is the process of understanding an unobserved phenomenon on the basis of a set of observations. For example, we may be interested in understanding variations in the district vote for the Conservative, Labour, and Social Democratic parties in Britain in 1979. We presumably have some hypotheses to evaluate; however, what we actually observe is 650 district elections to the House of Commons in that year.

Naively, we might think that we were directly observing the electoral strength of the Conservatives by recording their share of the vote by district and their overall share of seats. But a certain degree of randomness or unpredictability is inherent in politics, as in all of social life and all of scientific inquiry.[7] Suppose that in a sudden fit of absent-mindedness (or in deference to social science) the British Parliament had agreed to elections every week during 1979 and suppose (counterfactually) that these elections were independent of one another. Even if the underlying support for the Conservatives remained constant, each weekly replication would not produce the same number of votes for each party in each district. The weather might change, epidemics might break out, vacations might be taken—all these occurrences would affect voter turnout and electoral results. Additionally, fortuitous events might happen in the international environment, or scandals might reach the mass media; even if these had no long-term significance, they could affect the weekly results. Thus, numerous, transitory events could effect slightly different sets of election returns. Our observation of any one election would not be a perfect measure of Conservative strength after all.

As another example, suppose we are interested in the degree of conflict between Israelis (police and residents) and Palestinians in communities on the Israeli-occupied West Bank of the Jordan River. Official reports by both sides seem suspect or are censored, so we decide to conduct our own study. Perhaps we can ascertain the general level of conflict in different communities by intensive interviews or participa-

[7] See Popper (1982) for a book-length defense of indeterminism.

tion in family or group events. If we do this for a week in each community, our conclusions about the level of conflict in each one will be a function in part of whatever chance events occur the week we happen to visit. Even if we conduct the study over a year, we still will not perfectly know the true level of conflict, even though our uncertainty about it will drop.

In these examples, the variance in the Conservative vote across districts or the variance in conflict between West Bank communities can be conceptualized as arising from two separate factors: *systematic* and *nonsystematic* differences. Systematic differences in our voter example include fundamental and predictable characteristics of the districts, such as differences in ideology, in income, in campaign organization, or in traditional support for each of the parties. In hypothetical weekly replications of the same elections, systematic differences would persist, but the nonsytematic differences such as turnout variations due to the weather, would vary. In our West Bank example, systematic differences would include the deep cultural differences between Israelis and Palestinians, mutual knowledge of each other, and geographic patterns of residential housing segregation. If we could start our observation week a dozen different times, these systematic differences between communities would continue to affect the observed level of conflict. However, nonsystematic differences, such as terrorist incidents or instances of Israeli police brutality, would not be predictable and would only affect the week in which they happened to occur. With appropriate inferential techniques, we can usually learn about the nature of systematic differences even with the ambiguity that occurs in one set of real data due to nonsystematic, or random, differences.

Thus, *one of the fundamental goals of inference is to distinguish the systematic component from the nonsystematic component of the phenomena we study*. The systematic component is not more important than the nonsystematic component, and our attention should not be focused on one to the exclusion of the other. However, distinguishing between the two is an essential task of social science. One way to think about inference is to regard the data set we compile as only one of many possible data sets—just as the actual 1979 British election returns constitute only one of many possible sets of results for different hypothetical days on which elections could have been held, or just as our one week of observation in one small community is one of many possible weeks.

In descriptive inference, we seek to understand the degree to which our observations reflect either typical phenomena or outliers. Had the 1979 British elections occurred during a flu epidemic that swept through working-class houses but tended to spare the rich, our observations might be rather poor measures of underlying Conservative

strength, precisely because the nonsystematic, chance element in the data would tend to overwhelm or distort the systematic element. If our observation week had occurred immediately after the Israeli invasion of Southern Lebanon, we would similarly not expect results that are indicative of what usually happens on the West Bank.

The political world is theoretically capable of producing multiple data sets for every problem but does not always follow the needs of social scientists. We are usually only fortunate enough to observe one set of data. For purposes of a model, we will let this one set of data be represented by one variable y (say, the vote for Labor) measured over all $n = 650$ units (districts): $y_1, y_2, \ldots, y_n$ (for example, y_1 might be 23,562 people voting for Labor in district 1). The set of *observations* which we label y is a *realized variable*. Its values vary over the n units. In addition, we define Y as a *random variable* because it varies randomly across hypothetical replications of the same election. Thus, y_5 is the number of people voting for Labor in district 5, and Y_5 is the random variable representing the vote across many hypothetical elections that could have been held in district 5 under essentially the same conditions. The observed votes for the Labor party in the one sample we observe, $y_1, y_2, \ldots, y_n$, differ across constituencies because of systematic and random factors. That is, to distinguish the two forms of "variables," we often use the term *realized variable* to refer to y and *random variable* to refer to Y.

The same arrangement applies to our qualitative example. We would have no hope or desire of quantifying the level of tension between Israelis and Palestinians, in part because "conflict" is a complicated issue that involves the feelings of numerous individuals, organizational oppositions, ideological conflicts, and many other features. In this situation, y_5 is a realized variable which stands for the total conflict observed during our week in the fifth community, say El-Bireh.[8] The random variable Y_5 represents both what we observe in El-Bireh and what we could have observed; the randomness comes from the variation in chance events over the possible weeks we could have chosen to observe.[9]

One goal of inference is to learn about *systematic features* of the random variables $Y_1, \ldots, Y_n$. (Note the contradictory, but standard, terminology: although in general we wish to distinguish systematic from nonsystematic components in our data, in a specific case we wish to

[8] Obviously the same applies to all the other communities we might study.

[9] Note that the randomness is not exactly over different actual weeks, since both chance events and systematic differences might account for observed differences. We therefore create the more ideal situation in which we imagine running the world again with systematic features held constant and chance factors allowed to vary.

take a random variable and extract its systematic features.) For example, we might wish to know the expected value of the Labor vote in district 5 (the average Labor vote Y_5 across a large number of hypothetical elections in this district). Since this is a systematic feature of the underlying electoral system, the expected value is of considerable interest to social scientists. In contrast, the Labor vote in one observed election, y_5, is of considerably less long-term interest since it is a function of systematic features *and* random error.[10]

The expected value (one feature of the systematic component) in the fifth West Bank community, El-Bireh, is expressed formally as follows:

$$E(Y_5) = \mu_5$$

where $E(\cdot)$ is the expected value operation, producing the average across an infinite number of hypothetical replications of the week we observe in community 5, El-Bireh. The parameter μ_5 (the Greek letter mu with a subscript 5) represents the answer to the expected value calculation (a level of conflict between Palestinians and Israelis) for community 5. This parameter is part of our model for a systematic feature of the random variable Y_5. One might use the observed level of conflict, y_5, as an estimate of μ_5, but because y_5 contains many chance elements along with information about this systematic feature, better estimators usually exist (see section 2.7).

Another systematic feature of these random variables which we might wish to know is the level of conflict in the *average* West Bank community:

$$\frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_i = \mu \tag{2.2}$$

One estimator of μ might be the average of the observed levels of conflict across all the communities studied, $\bar{y}$, but other estimators for this systematic feature exist, too. (Note that the same summary of data in our discussion of summarizing historical detail from section 2.5 is used for the purpose of estimating a descriptive inference.) Other systematic features of the random variables include the variance and a variety of causal parameters introduced in section 3.1.

Still another systematic feature of these random variables that might be of interest is the variation in the level of conflict within a community even when the systematic features do not change: the extent to which observations over different weeks (different hypothetical realizations of the same random variable) produce divergent results. This is, in other words, the size of the nonsystematic component. Formally, this is calculated for a single community by using the variance (instead of the expectation):

$$V(Y_i) = \sigma_i^2 \tag{2.3}$$

where σ^2 (the Greek letter sigma) denotes the result of applying the variance operator to the random variable Y_i. Living in a West Bank community with a high level of conflict between Israelis and Palestinians would not be pleasant, but living in a community with a high variance, and thus unpredictability, might be worse. In any event, both may be of considerable interest for scholarly researchers.

To understand these issues better, we distinguish two fundamental views of random variation.[11] These two perspectives are extremes on a continuum. Although significant numbers of scholars can be found who are comfortable with each extreme, most political scientists have views somewhere between the two.

Perspective 1: A Probabilistic World. Random variation exists in nature and the social and political worlds and can never be eliminated. Even if we measured all variables without error, collected a census (rather than only a sample) of data, and included every conceivable explanatory variable, our analyses would still never generate perfect predictions. A researcher can divide the world into apparently systematic and apparently nonsystematic components and often improve on predictions, but nothing a researcher does to analyze data can have any effect on reducing the fundamental amount of nonsystematic variation existing in various parts of the empirical world.

Perspective 2: A Deterministic World. Random variation is only that portion of the world for which we have no explanation. The division between systematic and stochastic variation is *imposed* by the analyst and depends on what explanatory variables are available and included in the analysis. Given the right explanatory variables, the world is entirely predictable.

These differing perspectives produce various ambiguities in the inferences in different fields of inquiry.[12] However, for most purposes

[10] Of course, y_5 may be of tremendous interest to the people in district 5 for that year, and thus both the random and systematic components of this event might be worth studying. Nevertheless, we should always try to distinguish the random from the systematic.

[11] See King (1991b) for an elaboration of this distinction.

[12] Economists tend to be closer to Perspective 1, whereas statisticians are closer to Perspective 2. Perspective 1 is also especially common in the field of engineering called "quality control." Physicists have even debated this distinction in the field of quantum mechanics. Early proponents of Perspective 2 subscribed to the "hidden variable theory"

these *two perspectives can be regarded as observationally equivalent*. This is especially true if we assume, under Perspective 2, that at least some explanatory variables remain unknown. Thus, observational equivalence occurs when these unknown explanatory variables in Perspective 2 become the interpretation for the random variation in Perspective 1. Because of the lack of any observable implications with which to distinguish between them, a choice between the two perspectives depends on faith or belief rather than on empirical verification.

As another example, with both perspectives, distinguishing whether a particular political or social event is the result of a systematic or nonsystematic process depends upon the choices of the researcher. From the point of view of Perspective 1, we may tentatively classify an effect as systematic or nonsystematic. But unless we can find another set of data (or even just another case) to check for the persistence of an effect or pattern, it is very difficult to make the right judgment.

From the extreme version of Perspective 2, we can do no more than describe the data—"incorrectly" judging an event as stochastic or systematic is impossible or irrelevant. A more realistic version of this perspective admits to Perspective 1's correct or incorrect attribution of a pattern as random or systematic, but it allows us some latitude in deciding what will be subject to examination in any particular study and what will remain unexplained. In this way, we begin any analysis with all observations being the result of "nonsystematic" forces. Our job is then to provide evidence that particular events or processes are the result of systematic forces. Whether an unexplained event or process is a truly random occurrence or just the result of as yet unidentified explanatory variables is left as a subject for future research.

This argument applies with equal force to qualitative and quantitative researchers. Qualitative research is often historical, but it is of most use as social science when it is also explicitly inferential. To conceptualize the random variables from which observations are generated and to attempt to estimate their systematic features—rather than merely summarizing the historical detail—does *not* require large-scale data collections. Indeed, one mark of a good historian is the ability to distinguish systematic aspects of the situation being described from idiosyncratic ones. This argument for descriptive inference, therefore, is certainly not a criticism of case studies or historical work. Instead,

of quantum mechanics. However, more modern work seems to provide a fundamental verification of Perspective 1: the physical world seems intrinsically probabilistic. We all await the resolution of the numerous remaining contradictions of this important theory and its implications for the nature of the physical world. However, this dispute in physics, although used to justify much of the philosophy of social science, is unlikely to affect the logic of inference or practice of research in the social sciences.

any kind of social science research should satisfy the basic principles of inference discussed in this book. Finding evidence of systematic features will be more difficult with some kinds of evidence, but it is no less important.

As an example of problems of descriptive inference in historical research, suppose that we are interested in the outcomes of U.S.–Soviet summit meetings between 1955 and 1990. Our ultimate purpose is to answer a causal question: under what conditions and to what extent did the summits lead to increased cooperation? Answering that question requires resolving a number of difficult issues of causal analysis, particularly those involving the direction of causality among a set of systematically related variables.[13] In this section, however, we restrict ourselves to problems of descriptive inference.

Let us suppose that we have devised a way of assessing—through historical analysis, surveying experts, counting "cooperative" and "conflictual" events or a combination of these measurement techniques—the extent to which summits were followed by increased superpower cooperation. And we have some hypotheses about the conditions for increased cooperation—conditions that concern shifts in power, electoral cycles in the United States, economic conditions in each country, and the extent to which previous expectations on both sides have been fulfilled. Suppose also that we hope to explain the underlying level of cooperation in each year, and to associate it somehow with the presence or absence of a summit meeting in the previous period, as well as with our other explanatory factors.

What we observe (even if our indices of cooperation are perfect) is only the degree of cooperation *actually* occurring in each year. If we observe high levels of cooperation in years following summit meetings, we do not know without further study whether the summits and subsequent cooperation are systematically related to one another. With a small number of observations, it could be that the association between summits and cooperation reflects randomness due to fundamental uncertainty (good or bad luck under Perspective 1) or to as yet unidentified explanatory variables (under Perspective 2). Examples of such unidentified explanatory variables include weather fluctuations leading to crop failures in the Soviet Union, shifts in the military balance, or leadership changes, all of which could account for changes in the extent of cooperation. If identified, these variables are alternative explanations—omitted variables that could be collected or examined

[13] In our language, as we will discuss in section 3.5 below, the issue is that of *endogeneity*. Anticipated cooperation could lead to the convening of summit meetings, in which case, instead of summit meetings explaining cooperation, anticipated cooperation would explain actual cooperation—hardly a startling finding if actors are rational!

to assess their influence on the summit outcome. If unidentified, these variables may be treated as nonsystematic events that could account for the observed high degree of superpower cooperation. To provide evidence against the possibility that random events (unidentified explanatory variables) account for the observed cooperation, we might look at many other years. Since random events and processes are by definition not persistent, they will be extremely unlikely to produce differential cooperation in years with and without superpower summits. Once again, we are led to the conclusion that only repeated tests in different contexts (years, in this case) enable us to decide whether to define a pattern as systematic or just due to the transient consequences of random processes.

Distinguishing systematic from nonsystematic processes is often difficult. From the perspective of social science, a flu epidemic that strikes working-class voters more heavily than middle-class ones is an unpredictable (nonsystematic) event that in one hypothetical replication of the 1979 election would decrease the Labor vote. But a persistent pattern of class differences in the incidence of a disabling illness would be a systematic effect lowering the average level of Labor voting across many replications.

The victory of one candidate over another in a U.S. election on the basis of the victor's personality or an accidental slip of the tongue during a televised debate might be a random factor that could have affected the likelihood of cooperation between the USSR and the United States during the Cold War. But if the most effective campaign appeal to voters had been the promise of reduced tensions with the USSR, consistent victories of conciliatory candidates would have constituted a systematic factor explaining the likelihood of cooperation.

Systematic factors are persistent and have consistent consequences when the factors take a particular value. Nonsystematic factors are transitory: we cannot predict their impact. But this does not mean that systematic factors represent constants. Campaign appeals may be a systematic factor in explaining voting behavior, but that fact does not mean that campaign appeals themselves do not change. It is the *effect* of campaign appeals on an election outcome that is constant—or, if it is variable, it is changing in a predictable way. When Soviet-American relations were good, promises of conciliatory policies may have won votes in U.S. elections; when relations were bad, the reverse may have been true. Similarly, the weather can be a random factor (if intermittent and unpredictable shocks have unpredictable consequences) or a systematic feature (if bad weather always leads to fewer votes for candidates favoring conciliatory policies).

In short, summarizing historical detail is an important intermediate

step in the process of using our data, but we must also make descriptive inferences distinguishing between random and systematic phenomena. Knowing what happened on a given occasion is not sufficient by itself. *If we make no effort to extract the systematic features of a subject, the lessons of history will be lost, and we will learn nothing about what aspects of our subject are likely to persist or to be relevant to future events or studies.*

2.7 Criteria for Judging Descriptive Inferences

In this final section, we introduce three explicit criteria that are commonly used in statistics for judging methods of making inferences—unbiasedness, efficiency, and consistency. Each relies on the random-variable framework introduced in section 2.6 but has direct and powerful implications for evaluating and improving qualitative research. To clarify these concepts, we provide only the simplest possible examples in this section, all from descriptive inference. A simple version of inference involves estimating parameters, including the expected value or variance of a random variable (μ or σ^2) for a descriptive inference. We also use these same criteria for judging causal inferences in the next chapter (see section 3.4). We save for later chapters specific advice about doing qualitative research that is implied by these criteria and focus on the concepts alone for the remainder of this section.

2.7.1 *Unbiased Inferences*

If we apply a method of inference again and again, we will get estimates that are sometimes too large and sometimes too small. Across a large number of applications, do we get the right answer *on average*? If yes, then this method, or "estimator," is said to be unbiased. This property of an estimator says nothing about how far removed from the average any one application of the method might be, but being correct on average is desirable.

Unbiased estimates occur when the variation from one replication of a measure to the next is nonsystematic and moves the estimate sometimes one way, sometimes the other. Bias occurs when there is a systematic error in the measure that shifts the estimate more in one direction than another over a set of replications. If in our study of conflict in West Bank communities, leaders had created conflict in order to influence the study's results (perhaps to further their political goals), then the level of conflict we observe in every community would be biased toward greater conflict, on average. If the replications of our