How to Think about Statistics

Fifth Edition

John L. Phillips, Jr.
Boise State University

1

Introduction

You may be planning to study statistics not because you want to but because you have to. If so, I know how you feel. I went through the same experience years ago; if I could have avoided statistics, I probably would have. However, my attitude changed after I began to study it, for I discovered in it a new way of thinking that was truly fascinating.

But your present task may be even more challenging than mine was. You won't have to do the computations that I did, but you are about to acquire within a very short time (and possibly by yourself) the same grasp of the underlying structure of statistics that I acquired in two full semesters under an excellent teacher.

The Task

Your plight and your prospects are well illustrated, I think, by the experience of a student who was asked to evaluate the prototype of this book:

> It was the most difficult book I have ever read. It was foreign to me since I had had no background knowledge of the things talked about. . . . If I was to understand it, I realized I would need to outline the book chapter by chapter. I did so, and to my amazement, it followed a very orderly pattern after all. It really did present what the author had stated he hoped to present. If you understood Chapter 1, you could

This quotation contains some important advice on how to use the book. I would only add that even though you may have mastered the ideas preceding the one you are working on at any given moment, you should be prepared to go back to those ideas from time to time and consider their relation to the new one being presented. I have tried, by providing frequent cross-references, to help you do just that. (You may wish to keep a couple of bookmarks handy for that purpose.) When you are finished, you should have in mind a hierarchical structure, with each new idea related to one or more of the ideas that have preceded it.

The conceptualization of such a structure is highly satisfying in itself, but there are many other reasons for making the effort. It is true that an understanding of statistical concepts will not be of critical assistance to you in gathering economic data, in conducting interviews for a political or sociological opinion or attitude study, in uncovering archeological artifacts, or in teaching children. But often people who do economic, political, sociological, anthropological, archeological, or educational studies (to name but a few) report their findings in statistical terms. If you are planning a career in any of the several professions to which those studies are relevant, it is important that you be able to read them with understanding. A continuing awareness of developments in one's field is the mark of the true professional, and this book will help you maintain that awareness.

THE BASIC IDEAS

To understand the meaning of any measurement in the social sciences, you must know at least two things about it. First, you must be able to describe the operations by which it was obtained, and second, you must be able to compare it with other measurements that have been obtained in the same way.

This book is concerned primarily with the second kind of knowledge. Statistical thinking deals with multiple measurements. It analyzes the relation of how many to how much—of frequencies to scores. If the basic element in measurement itself is a *score*, the corresponding concept in statistics is a *distribution* that includes many scores—in short, a *frequency distribution.*

Such a distribution can be described by drawing a picture of it—and indeed in the early stages of your learning about distributions, that is the method I shall use to describe them. The method is cumbersome, however; so ways have been devised to achieve roughly the same result through the use of numbers rather than diagrams. The most important advantage of numbers over diagrams is that they can be manipulated in a way that diagrams cannot.

Imagine that (for reasons known only to your psychoanalyst) you have just had a pile of gravel dumped on your front lawn and that (for similar reasons) you want to describe the result to me over the telephone. To do that successfully, you will

0026-3

that is, whether it is shaped like a cone, a pancake, or perhaps your garage roof; (2) its *location*—that is, how far and in what direction it is from some reference point familiar to me; and (3) its *dispersion*—the extent to which it is spread out—for example, if it is a cone, is it a steep-sided one that covers only a small area, or is it a low-profile one that covers most of the lawn?

That pile of pebbles is analogous to a *frequency distribution* of scores, and the same kinds of information are needed to describe either adequately. Concerning configuration, we have certain names for frequency distributions that convey information to anyone who is familiar with them—names such as "normal," "symmetrical," "positively skewed," and "bimodal." Concerning location, the procedure is pretty much the same as it is for a pile of gravel: A dimension is identified and a reference point is chosen, and measurement of distance to the center of the distribution is from that reference point. The result is a measure of *central tendency.* Finally, to communicate information about the amount of dispersion, a new reference point is used—namely, the center of the distribution—and the needed information may consist of the average distance of individuals (like that of the individual pebbles in a pile of gravel) from the central point. That distance is a measure of *variability* (dispersion).

But just describing a distribution is not always enough. Often you will be interested in *two* distributions and in the relationship that exists between them. Consider a single variable, "IQ in the general population," and a few other variables with which it might be paired: family income, some index of health care, a general index of socioeconomic status, race, or place of residence. Other interesting relationships could be investigated among the IQs of various subgroups of the general population: between parents and their children, between identical twins, between fraternal twins, between non-twin siblings, and between pairs of unrelated children. These are just a few of the relationships that come to mind at the moment. Others would occur to you if you were making a study of intelligence and its correlates, and in every case you would need a way of communicating your findings to others; in short, you would need a measure of *correlation.*

Whether or not you wish to relate one set of scores to another, you will surely want to be able to tell what each individual measure means. If you are a teacher who has given my child a test, and I ask you how well he did on it, you might try to put me off by saying his score was "high" or "low" and go on to talk about something else. But if I want to know *how* high, you are in trouble. You may answer that he got 90 percent of the test items right. You think you are off the hook, but I persist: "How *hard* is that test? Ninety percent is very impressive if the items are all difficult, but not if most of the other kids score above 95!"

Our conversation has been concerned entirely with the *interpretation of individual measures*, and I'm sure you'll agree that all of my questions are pertinent. Without answers to them and others like them, one really cannot know the meaning

On the other hand, you must be careful to avoid overinterpreting measures, whether they be of individuals or of groups. If, for example, you were to weigh a *random sample* *,† of fifty 10-year-old boys, could you easily compute a measure of central tendency for the sample that is identical to the central tendency of a population that includes *all* 10-year-old boys? How different might the obtained weight be if it were computed from a different random sample of that same population? Questions like these have to do with *precision of inference*, which is a kind of *reliability*, and we do have ways of dealing with them.

Assuming for the moment that you do know how to deal with questions about precision, consider this one additional question: Upon further analysis of your sample of 10-year-old boys, you discover a marked difference between the weights of boys living in one area of the country and those living in another. You suspect that the difference is due to diet, and you subsequently narrow that hypothesis down to a single vitamin that seems to be more plentiful in one of the areas than in the other. One way to test your hypothesis would be to select two samples of male infants who are living in the area in which the vitamin is less plentiful, introduce the suspected vitamin into the diet of one group, and then after a period of nine years weigh the subjects in both samples again. Let's imagine that there is a difference between the two. Is it large enough that you can be reasonably sure that it did not occur by chance—that a replication of the same study would not turn up a difference of zero, or even a difference in the other direction? To put it another way, how *significant* is the difference you obtained? Again, there are ways of dealing with such questions.

In the chapters that follow, each of the above ideas will be developed further, but always in the manner that you have seen here. The discussion is aimed directly at the underlying *logic* of statistical thinking, with an absolute minimum of arithmetical and algebraic manipulations. You will find the logic similar in many ways to common sense. The main difference is that the logic presented here is rigorously systematic, and like any system, its parts are interdependent. So the book cannot be studied piecemeal; the ordering of the chapters is deliberate and necessary. Once

* A random sample is one in which (1) every member of the population has an equal chance of being included in the sample and (2) each selection is made independently of all the others.

† The notes in this book are of two kinds and are denoted by two kinds of superscript. Footnotes are indicated by conventional footnote symbols (*, †, etc.); other notes, which appear in the back of the book, are indicated by numbers.

The footnotes are intended to supplement the central discussion at any given time and are important to a full understanding of that discussion. A footnote may refine an idea by restricting or extending its implication; it may explain a pedagogical technique by commenting on the relative importance of its various attributes; or it may supply cross-references that are not essential but are enriching. In fact, the objective in every case is enrichment. The way to use the footnotes, then, is (1) *to ignore them* as you work through a section for the first time and (2) *to study them carefully* the second time through. Every chapter has a few footnotes.

0026-4

you have worked through it, however, the book will serve as a convenient reference for you in your professional life. It has been organized with that in mind.

Description of Data versus Inference to Population

A reader trained in statistics would have noticed in the preceding section a subtle change between the eighth and ninth paragraphs. Every comment before that point was concerned with the description of a set of data—its configuration, average value, dispersion, and relation to other data sets. But when we began to consider taking a sample from a population and estimating properties of that population from what we know about the sample, we were shifting from description to inference.

That is an important shift—so important that I have devoted a special chapter to it. Chapter 7 is entitled "Description to Inference: A Transition."

Facing Mathphobia

Many persons who are intelligent and who perform many other tasks well find themselves frozen in fear when confronted by any mathematical problem beyond the level of basic arithmetic. If you are not afflicted with such mathphobia, skip this entire section and proceed to Chapter 2. If you do have such a phobia, I am not going to attempt to rid you of it; no book is likely to accomplish that. What I believe I *can* do in this section is show you that your phobia need not be aroused by the contents of this volume, because there is very little mathematics here beyond basic arithmetic—that is, the four fundamental processes of addition, subtraction, multiplication, and division. By "very little" I mean (1) formulas, (2) some arithmetic that is more advanced than the four fundamental processes, and (3) graphic representations of data.

Because most of the concepts in these three categories are or have at some time in the past been familiar to you, the remainder of this section is mostly a very brief review. The one concept that may be new to you (the frequency distribution) is probably the easiest of them all; nevertheless, because it is new, it is treated separately in Chapter 2.

Formulas

You may have thought of formulas as guides to computations: You plug numbers into a formula, follow the rules of algebra, and out comes an answer. Formulas can indeed be very useful in that way, but the emphasis in this book is on the *concepts*—the logical structures—that underlie the computations.

For that reason, the only formulas in the book are definitional. (A *definitional* there only to illustrate the concepts, and the concepts are illumined by the definitional formulas.

So when you see a new formula, try to discover what it means; look for the relations that it specifies. For our purposes, that specification usually need not be very precise; you may think, for example, of one term in an equation as being "larger" or "smaller" than another instead of "3.14 times as large" or "$\frac{1}{3.14}$th as large." Or you may note that as one term becomes larger, another becomes smaller. For our purposes, that frequently is the most important observation you can make. For example, take the equation

$$v = \frac{d}{t}$$

where v is velocity, d is distance, and t is time. It says that you can find v by dividing the numerator (d) of the expression d/t by the denominator (t). A body that moves a long distance in a given amount of time is moving faster than one that moves a short distance in the same amount of time. Conversely, one that takes a long time to cover a particular distance is moving *less* rapidly than one that takes a short time to cover the same distance. The two sides of any equation are equal by definition; so if there is a change in one side, the other must change, too, in a way that restores equilibrium. Thus, increasing a numerator in a right-hand term has the effect of *increasing* the left-hand term, while increasing the denominator makes the left term *smaller*. To understand basic relationships, this kind of knowledge is really all you need.

If, after you have analyzed a formula in this way, you want to pursue its implications for manipulating data, refer to the "calculation box" that you will find near the text where the formula is introduced. If you want to investigate still further, consult a text on statistical methods. But calculation is not our main concern; for us, it is but another way of illustrating a concept. Our major concern is the comprehension of the concept rather than the calculation of an answer that is numerically correct.

Arithmetic

Concerning the issue of more advanced arithmetic, three concepts will suffice for comprehending the ideas presented in this book: (1) the square, (2) the square root, and (3) negative numbers. If you do suffer from mathphobia but do not have any difficulty with these three concepts, skip the rest of this subsection.

The *square* of a given number is that number multiplied by itself. If you square 3, you multiply 3 by itself; the result is 9.*

* Probably the most common symbol for multiplication is an x interposed between the values that are to be multiplied. When x is also being used to represent a *value*, however, using it to signify an operation (multiplication) can be confusing. One way to avoid the confusion is to let *parentheses* signify multiplication—that is, any value bounded by parentheses is to be multiplied by whatever value is represented immediately adjacent to either parenthesis. Thus 2(2) = 2 × 2 = 4 and 4(3 + 17 + 5) =

$$3^2 = 3 \text{ times } 3 = 3(3) = 9$$

$$10^2 = 10 \text{ times } 10 = 10(10) = 100$$

$$100^2 = 100 \text{ times } 100 = 100(100) = 10{,}000$$

The *square root* of a given number is the number that when multiplied by itself produces the given number. After you have squared a number (e.g., $3^2 = 9$), taking a square root of the result ($\sqrt{9} = 3$) gets you back to where you started, namely, 3.

$$3^2 = 3(3) = 9 \quad \text{so} \quad \sqrt{9} = 3$$

$$10^2 = 10(10) = 100 \quad \text{so} \quad \sqrt{100} = 10$$

$$100^2 = 100(100) = 10{,}000, \text{ so } \sqrt{10{,}000} = 100$$

Conversely, after you have taken the square root of a number (e.g., $\sqrt{9} = 3$), squaring the result gets you back to where you started (9).

$$\sqrt{9} = 3 \quad \text{so} \quad 3^2 = 3(3) = 9$$

$$\sqrt{100} = 10 \quad \text{so} \quad 10^2 = 10(10) = 100$$

$$\sqrt{10{,}000} = 100 \text{ so } 100^2 = 100(100) = 10{,}000$$

In general, since the operations of squaring and taking the square root are precisely the inverse of each other, squaring the square root of a quantity yields the quantity that you started with:

$$(\sqrt{x})^2 = (\sqrt{x})(\sqrt{x}) = x$$

Similarly, taking the square root of the square of a quantity also yields the quantity that you started with:*

$$\sqrt{x^2} = \sqrt{xx} = x$$

To put it another way, *the square of $\sqrt{x}$ is x, and the square root of x^2 is also x.*

A *negative number* is opposite in sign to a positive number. When you add such a number to a positive number of the same magnitude, the result is 0. If in Figure 1-1 you think of the horizontal line (the X-axis) as a balance beam with its fulcrum at 0, you can see that a −2 (i.e., negative 2) will balance a +2 (i.e., positive 2), a −3 will balance a +3, and so on. If you turn the page clockwise 90 degrees, the same will be true of the other axis (the Y-axis).

* When the terms that constitute an expression are *letters* rather than numbers, the multiplying operation

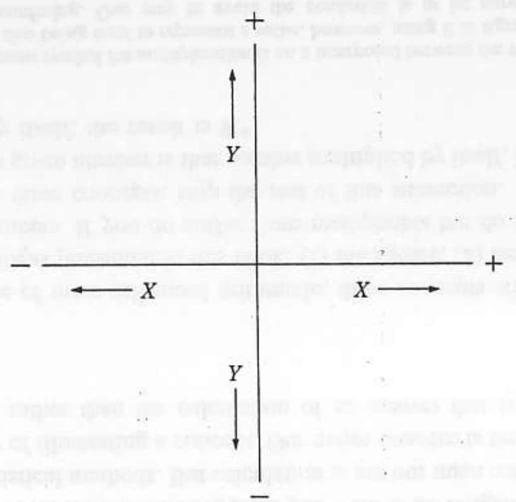


FIGURE 1-1 Two measures can be represented on a two-dimensional surface by a single point (often called a *data point*). If both of the measures are 0, that point will be precisely where the axis lines cross at the center of the graph.

If the data to be represented do not include any negative numbers—a frequent occurrence—a graph will consist of only the upper right-hand quadrant of Figure 1-1. (That's the part above the horizontal line and to the right of the vertical line.) But if there are negative numbers in the data, a surface similar to Figure 1-1 will be necessary to represent all the data graphically.

Graphs

Figure 1-1 lays the foundation for the remainder of this chapter, which is concerned entirely with *graphs*. Read the following descriptions rather quickly now; then return to them after you have finished the chapter.

You will find two kinds of graphs in this book:

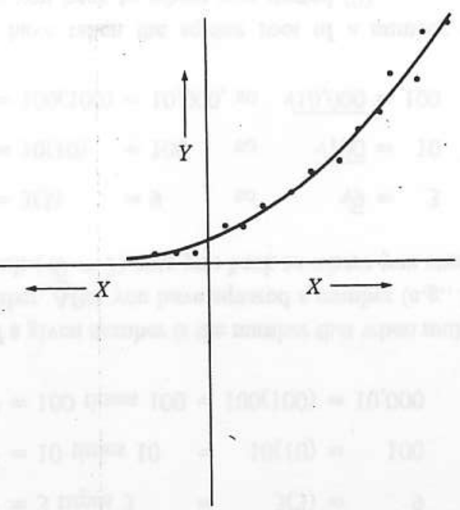
1. When the relation between two quantities (variables*) is plotted, the horizontal axis (X) represents changes in one of these quantities, and the vertical axis (Y) represents changes in the other. This is the classical meaning of the term *graph*.
2. When many objects are measured on one dimension X, the numbers of objects at various values of X can be represented as a stack of units at each of the many points of the X-axis. In this usage, Y is the height of the stack at each of those points. This is called a *frequency distribution* (see Chapter 2).

Some mathematicians prefer to reserve the term *graph* for the first of these two kinds of representation, but we shall accept a broader (and more common) meaning and refer to both kinds as graphs.

An example of the first kind of graph is one that represents the relationship between X, the temperature of the air in a room, and Y, the amount of moisture the air will hold—that is, how much water can be added before some of it condenses out of the air. The curve looks something like the one in Figure 1-2. Figure 1-3 shows how performance on a task varies with the performer's level of arousal; there is no zero on either axis. In each case, a dot represents a data point, and the simple curve that has been drawn is the one that best fits the points.

The second of the two kinds of graph described above is the frequency distribution. It is the focus of our next chapter.

But before we leave the first kind of graph, I want to warn you against drawing precipitous conclusions from your reading of *any* kind of graph. The two arrows in Figures 1-2 and 1-3 indicate the directions in which two variables, X and Y, *increase*. It is possible, however, for the larger magnitudes of a variable to be represented by the smaller number on the graph. If, for example, the Y variable is *skill at golf*, the lower scores (strokes per 18 holes) denote greater skill than do the higher ones. In a case like that, increase in skill could be represented by a falling rather than a rising curve. (The "curve" *could* be a straight line.) For example, if the X variable were *amount of practice* (measured in hours), the graph would look



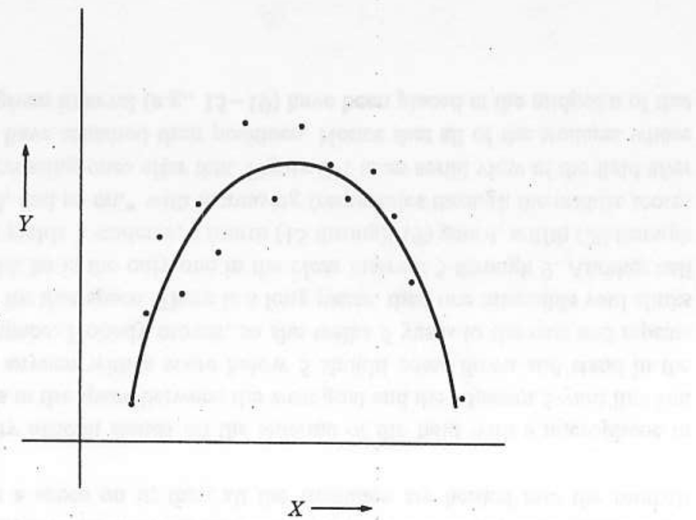
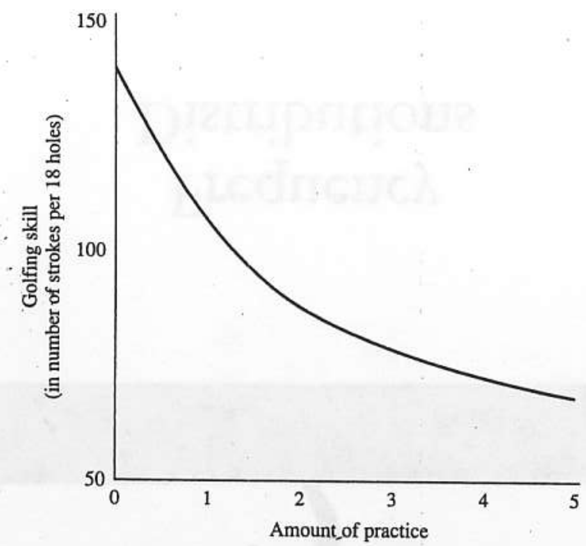


FIGURE 1-3 Graph showing the relation between level of arousal (X) and efficiency of performance (Y). There are no negative measures on either axis, and this graph uses only one quadrant of the surface displayed in Figure 1-1.



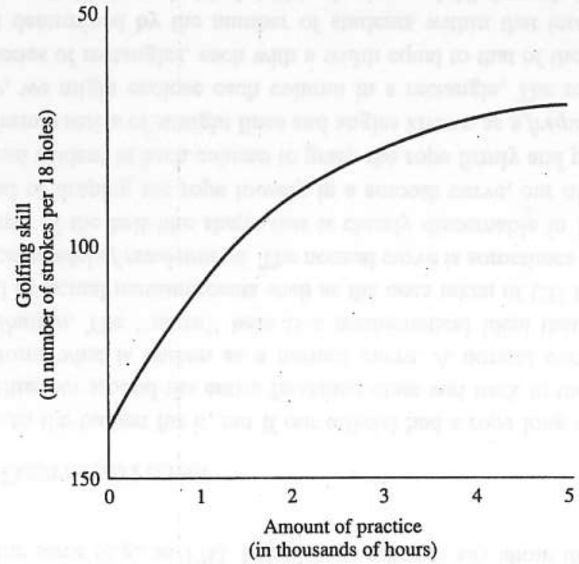


FIGURE 1-5 Graph of the relation between amount of practice (X) and golfing skill (Y). In order to show golfing skill rising instead of falling with an increasing amount of practice, the golf scores had to be plotted upside down.

something like Figure 1-4. The relation between X and Y is positive, but it appears to be negative because of the way Y is measured.

There is an alternative. To make the direction of change on the *graph* reflect that of the *Y variable* (in this case, golfing *skill*), the numbers on the Y-axis can be reversed. Figure 1-5 shows how that could be done with the golfing data. Figure 2-11, on page 22, illustrates a similar technique applied to *frequency* data. There, it is the numbers on the X-axis that have been reversed.

There is no generally accepted convention on this. Some writers are uncomfortable with a graph that even at first glance implies a relationship that is the opposite of the one intended. Others assume that every reader will carefully examine each axis of every graph and will infer only the relationship that is justified by that examination. So don't be satisfied with a first impression; it could be misleading.

2

Frequency Distributions

Every year, Central University administers a scholastic aptitude test to each of its incoming freshmen. After the tests have all been scored, each freshman is handed a small card with a score on it; then all the freshmen are herded into the football stadium.

A university official stands on the sideline of the field with a microphone in hand. She points to the space between the west goal and the adjacent 5-yard line and announces that anyone with a score below 5 should come down and stand in the middle of that space. Nobody moves, so she walks 5 yards to the east and repeats the instructions for that space. There is a long pause, then one miserable soul slinks down to the field; he is the only one in the *class interval* 5 through 9. Another call (10 through 14) yields 2 students, a fourth (15 through 19) gets 4, a fifth (20 through 24) produces 13, and so on,* with increasing frequencies through the middle scores followed by decreasing ones after that. Figure 2-1 is an aerial view of the field after all the students have assumed their positions. Notice that all of the students whose scores are in a given interval (e.g., 15–19) have been placed at the midpoint of that

* Technically these are all examples of the kind of class interval known as a *score interval*. The section

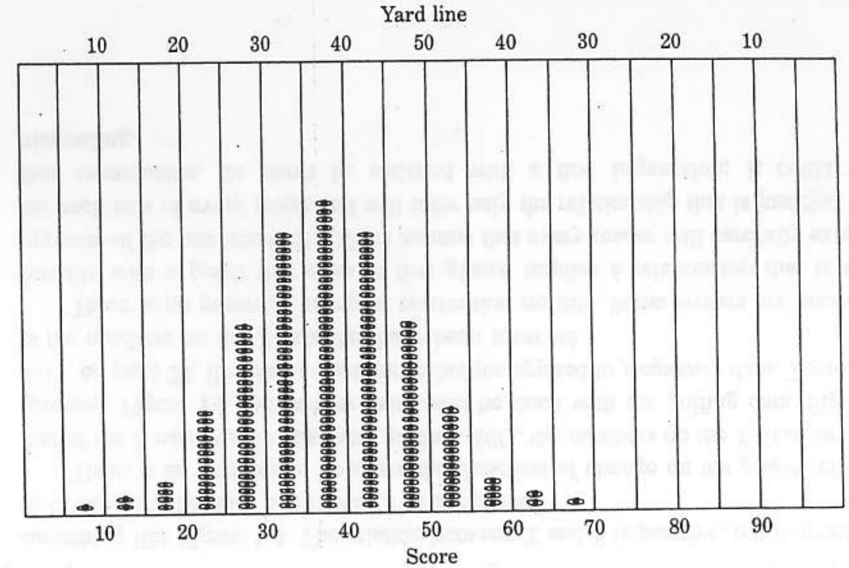


FIGURE 2-1 Students standing on football field in columns determined by scores on an aptitude test—a frequency distribution.

interval (e.g., four students at 17). For many purposes we may choose to treat all of those scores the same (e.g., as 17s). I shall have more to say about that very soon.

NORMAL DISTRIBUTIONS

Central U. lacks the budget for it, but if our official had a rope long enough to run from the sideline out around the entire freshman class and back to the sideline, the rope would form what is known as a *normal curve*. A normal curve encloses a *normal distribution*. The ''norm'' here is a mathematical ideal that is frequently approximated by actual measurements such as the ones taken of CU freshmen. It is *a mathematical model of randomness*. The normal curve is sometimes called a ''bell curve'' because of the bell-like shape that is clearly discernable in Figure 2-1.

If instead of draping the rope loosely in a smooth curve, our official were to instruct the end student in each column to grasp the rope firmly and pull it taut, the rope would form a series of straight lines and angles known as a *frequency polygon*. Alternatively, we might enclose each column in a rectangle. The resulting figure would be a series of rectangles, each with a width equal to that of the class interval and a height determined by the number of students within that interval—e.g., 2 within the interval 10 through 14, 4 within the interval 15 through 19, and so on.

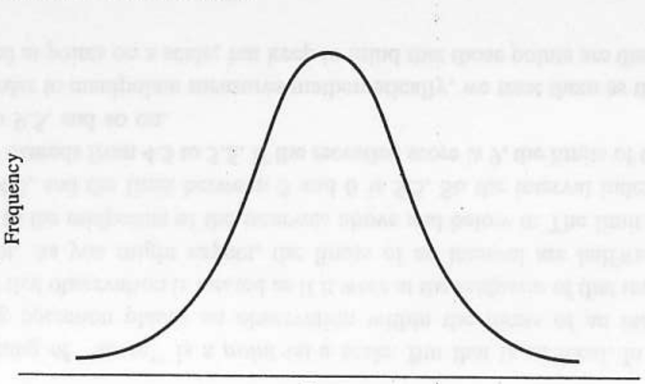
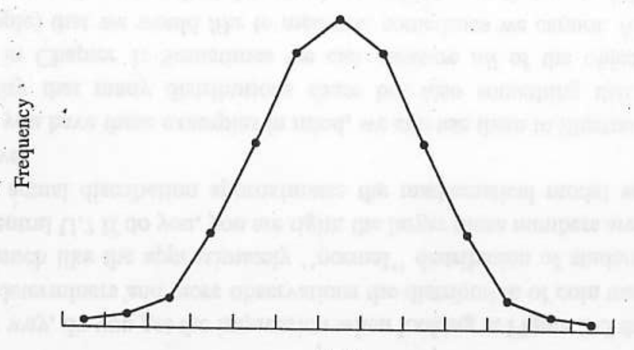


FIGURE 2-2 Normal curve from data of Figure 2-1.

recorded for each magnitude, the result is a frequency distribution. The *curve*, the *frequency polygon*, and the *histogram* are three ways of depicting a frequency distribution graphically (see Figures 2-2 through 2-4).

Quantities that are complexly determined do tend to form normal distributions. Scholastic aptitude measures show that tendency. Such measures have many determiners, both hereditary and environmental; some of those determiners influence an individual's score in an upward direction, some in a downward. In a few individuals, there is a preponderance of positive determiners; theirs are the scores in the upper (right-hand) *tail* of the distribution. In a few others, negative determiners predominate; their scores form the lower tail. Most scores, however, represent more balanced combinations of determiners; they form the large middle area of the distribution.



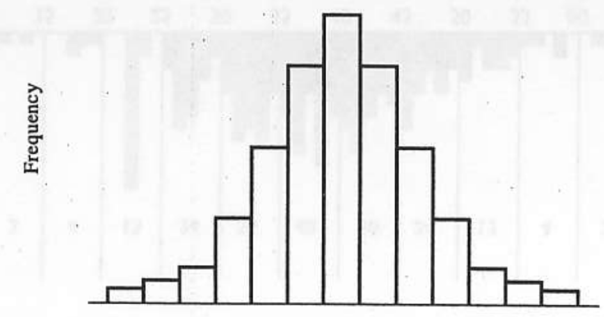
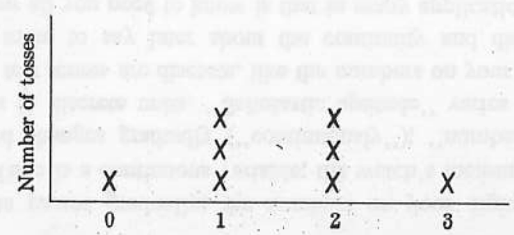


FIGURE 2-4 Histogram from data of Figure 2-1.

Each combination of determiners is assumed to be a matter of chance, and you can readily see that the probability of, say, 100 determiners being all positive (or all negative) in any individual would be vastly smaller than the probability of there being approximately half and half. If you *don't* readily see that, toss a mere three coins just eight times and plot the number of heads from each toss (0, 1, 2, and 3 are all possibilities).* You should come up with a distribution similar to the one in Figure 2-5 (although with such a small number of observations your distribution might differ markedly from that one). ''All heads'' (score 3) does not occur often by chance, and neither does all tails (score 0). The middle scores are much more frequent.

* In this example, the three coins correspond to 3 determiners (instead of the 100 cited in the example we just considered). The eight tosses correspond to 8 students instead of the 200 CU freshmen who took the aptitude test.



By the way, do you get the impression when looking at Figure 2-5 that if there were more determiners and more observations the distribution of coin tosses would look very much like the approximately "normal" distribution of students that we got from Central U.? If do you, you are right; the larger those numbers are, the more closely the actual distribution approximates the mathematical model we call the normal curve.

While you have these examples in mind, we can use them to illustrate not only the normality that many distributions share but also something that was first mentioned in Chapter 1: Sometimes we can measure *all* of the objects (for us usually people) that we would *like* to measure; sometimes we cannot. All of those objects taken together are called the *population* of interest; that part of the population that we study is called a *sample*.

The symbol for the size of a sample is *n*; for the *size* of the population it is *N*. In the case of the CU freshman, *n* is 200 if you intend to apply the results of your study to persons outside that group. If you do *not* plan to generalize, that particular class is the population of interest, and *N* is 200. If you *do* generalize beyond that group—say, to college freshman in general—you will be unable to count or measure the entire population; so anything you say about that population must be inferred from what you know about the sample.

The other example at hand is the coin experiment. There, *n* is 8, your inference is to *all* coin tosses, and again, *N* is unknown to you, as is everything else about the population. That is, you cannot observe and describe *all* coin tosses; you can only *infer* properties of such a population from those of the sample that you *do* observe. Our focus in this chapter is on description, not inference; I mention inference here because this section emphasizes normal distributions, and most of the inferential statistics in this book assume the normality of distributions.

GROUPED-FREQUENCY DISTRIBUTIONS AND THE MEANINGS OF SCORES

One meaning of "score" is a *point* on a scale. But that is an ideal. In practice a measuring operation places an observation within the limits of an *interval*, and thereafter that observation is treated as if it were at the *midpoint* of that interval even if it is not. As you might expect, the limits of an interval are halfway from its midpoint to the midpoints of the intervals above and below it: The limit between 5 and 4 is 4.5, and the limit between 5 and 6 is 5.5. So the interval indexed by the number 5 extends from 4.5 to 5.5. If the recorded score is 9, the limits of the interval are 8.5 to 9.5, and so on.

In order to manipulate measures mathematically, we treat them as though they are located at points on a scale; but keep in mind that those points are the midpoints

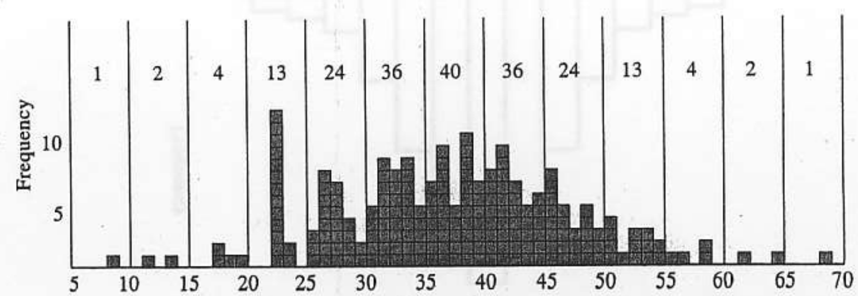
than *discrete*. Time passes gradually; the numbers on your digital wristwatch change abruptly. (Time is a continuous variable; the watch's measurement of it is not.) Driving speed changes gradually ("continuously"); "number of speeding tickets" progresses in discrete units. "Scholastic aptitude" varies continuously; scholastic aptitude test scores are discrete, like the numbers on your digital watch.

I shall have more to say later about the continuity and discontinuity of variables. Right now all you need to know is that in many applications, a reported score represents not a *point* on a scale of discrete units but an *interval* on a continuous scale and that it is nevertheless treated as though it were at the midpoint of that interval.

The conceptually simplest and at the same time most precisely accurate frequency distribution results from merely listing every score that is represented on the baseline and then counting the number of times each of them actually occurs. But doing it that way is like reconnoitering hilly terrain by hiking through its rocks and trees: You get a maximal amount of information, but it is not sufficiently organized to form the kind of pattern that you could see easily from an airplane.

Data generated by Central University's freshmen (Figure 2-1) have been organized into class intervals. To see how different the display might have looked if the data had *not* been "grouped," compare Figure 2-1 with Figure 2-6 and Figure 2-7A with Figure 2-7B. Figure 2-6 shows the entire 200 student scores ungrouped. In Figure 2-7A I have magnified a portion of Figure 2-3 (a frequency polygon that corresponds to the distribution in Figure 2-1) in order to emphasize the details exposed by the ungrouped configuration (Figure 2-7B).

If you count the number of students whose scores are in each of the class intervals of Figure 2-7A, you will find just 1 in the class interval 5 through 9. In the 10 through 14 interval there are 2; in the 15 through 19 category, 4; 20 through 24 holds 13 scores; and 25 through 29 has 24. Notice that corresponding intervals of



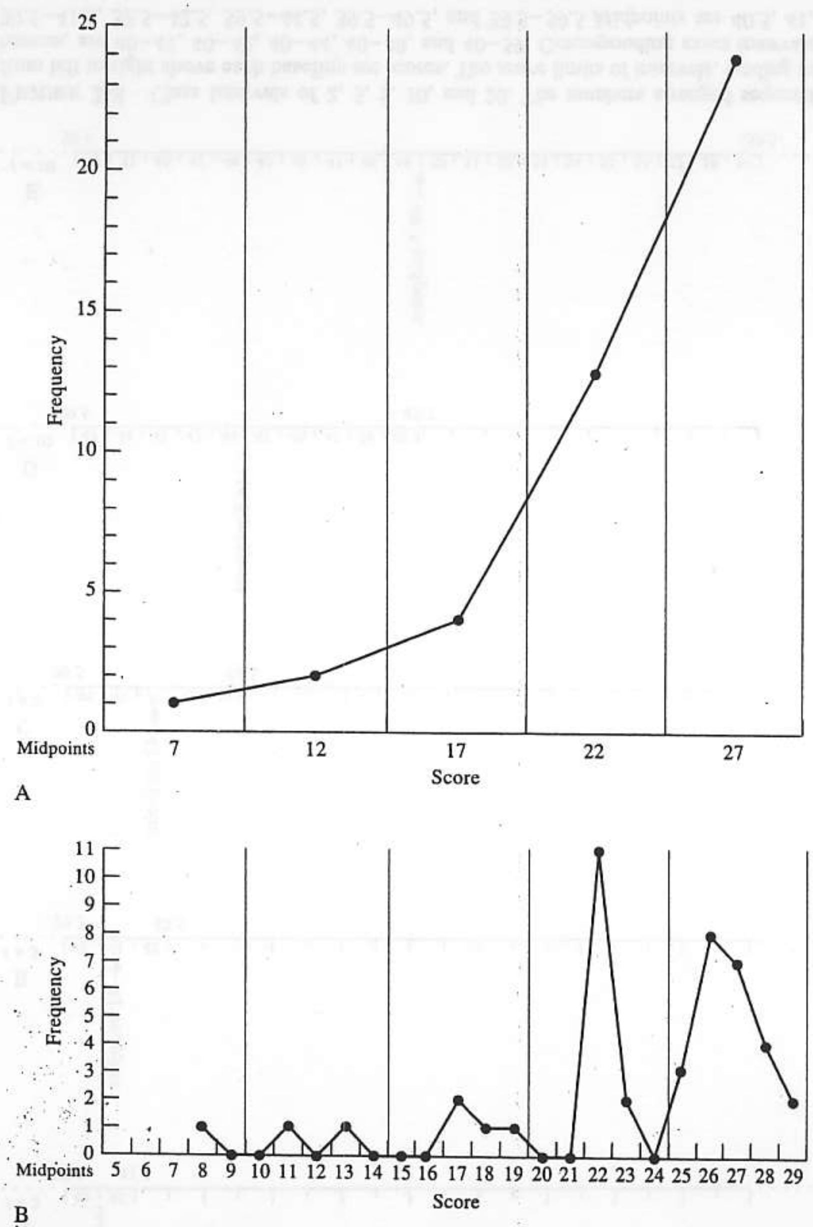


FIGURE 2-7(A) Magnification of the lower tail of the frequency polygon depicted in Figure 2-3, which corresponds to the distribution in Figure 2-1. **(B)** Same set of scores as in Figure 2-7A but without grouping.

Figure 2-7B contain exactly those same scores. The only difference is that in Figure 2-7A the scores have been "grouped"—i.e., *all* the scores in each interval have been moved to the midpoint of that interval, whereas Figure 2-7B gives you the precise location of each score. As you can see, that additional information obscures the general trend of the data.

You may have noticed also that the *numbers* denoting the scores 10, 20, and 30 have been shifted to the right of the *lines* (yard lines in Figure 2-1, mere ticks below the baseline in Figures 2-6, 2-7A, and 2-7B) that separate one class interval from another. The adjustment is precisely one-half of a unit. That is a refinement of Figure 2-1, because from the beginning I defined the lowest interval as "scores of 5 through 9," the next "10 through 14," and so on. Given that definition, the two limits of each score interval must be (1) below its lowest score and (2) above its highest score.

For example, a score of 10 is within the interval "10 through 14," *not* on the line *between* the intervals "5 through 9" and "10 through 14"; so the "10-yard line" in Figure 2-1 lies *below* a score of 10. That line is halfway between the highest score in the lower interval (in this case 9) and the lowest score in the higher interval (in this case 10); the line is therefore actually at 9.5. Accordingly, any score of 10 will be placed *above* the lower boundary of an interval that extends from 9.5 to 14.5, and all other intervals are marked off in the same way (e.g., 4.5 to 9.5 . . . 14.5 to 19.5 . . . 64.5 to 69.5). These are the exact class intervals that I promised on the first page of this chapter.

Five illustrations of the distinction between the score limits and the exact limits of a class interval are given in Figure 2-8. Score limits are inscribed above the baseline, exact limits below it.

This diagram also identifies (by means of a small arrow) the *midpoint* of each interval. Notice that the only midpoints that are whole numbers (41, 42) are in the two intervals (B and C) that are 3 and 5 score units long, respectively; 3 and 5 are odd numbers. Every interval that subtends an *even* number of units (2 in A, 10 in D, 20 in E) has a midpoint that falls *between* the middle *two* scores (40 and 41, 44 and 45, and 49 and 50, respectively). The midpoints of these three pairs of scores are 40.5, 44.5, and 49.5, none of which is a whole number. A close examination of Figure 2-8 should make it clear to you why that is the case. (To make it easy, just compare A and B, which contain only 2 and 3 score units, respectively.)

The importance of this discussion of midpoints lies in the fact that once data have been grouped into class intervals, the scores within each interval are treated as though they are all at the midpoint of that interval. For example, if you are working with a distribution of 30 scores grouped into the class intervals depicted in Figure 2-8A, subsequent calculations will be done with 30 scores none of which is a whole number. The same 30 scores deployed on the baseline pictured in Figure 2-8B will yield 30 whole numbers to be used in whatever calculations you have in mind.

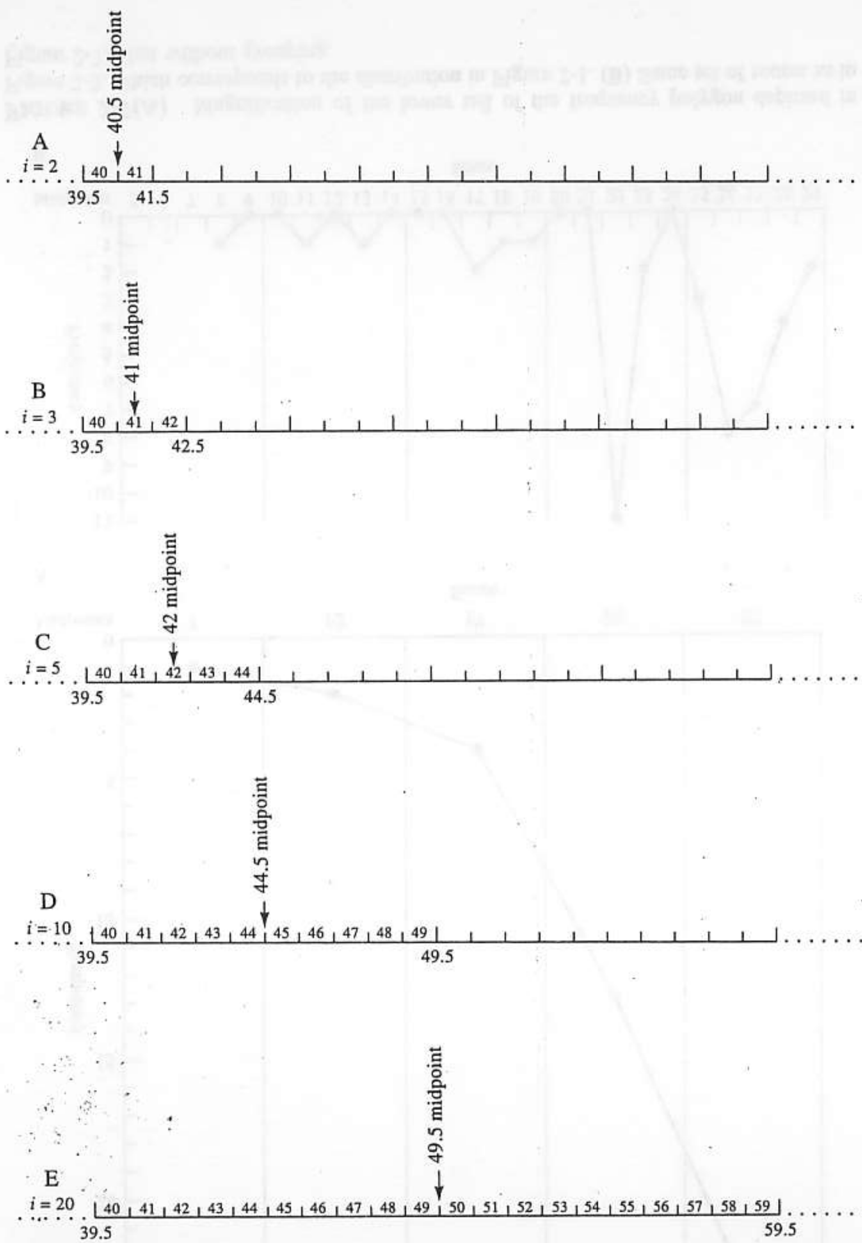


FIGURE 2-8 Class intervals of 2, 3, 5, 10, and 20. The numbers arranged sequentially from left to right above each baseline are scores. The *score* limits of intervals, reading top to bottom, are 40–41, 40–42, 40–44, 40–49, and 40–59. Corresponding exact intervals are 39.5–41.5, 39.5–42.5, 39.5–44.5, 39.5–49.5, and 39.5–59.5 *Midpoints* are 40.5, 41, 42,

0026-12

subtend odd numbers of score units have midpoints that are whole numbers, that kind of interval is frequently preferable to one containing an even number of units.*

SKEWED DISTRIBUTIONS

If the tail of a distribution is extended abnormally, that distribution is said to be *skewed.* When the direction of the extension is downward (i.e., toward the lower values), it is further categorized as *negatively* skewed. If your instructor in this course should give you an especially easy test, the distribution of your scores might look rather like Figure 2-9, which illustrates a negative skew.

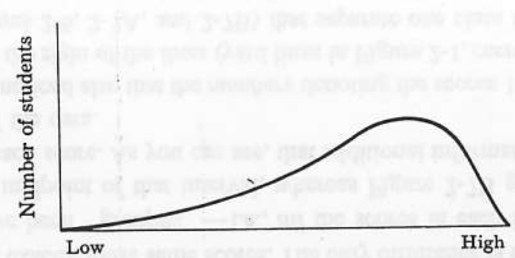
If the *upper* tail is similarly extended, the skew is said to be positive, and the distribution is *positively* skewed. A difficult test would produce a distribution more like Figure 2-10, because only a talented (or industrious) few deviate very far from the lowest possible score.

OTHER CONFIGURATIONS

There are other possible distributions. If we were measuring conformity behavior, like that of motorists at a busy intersection, we might get a J *curve*, like the one in Figure 2-11. The same would be true of a distribution of scores on a test that is extremely easy (so easy that most people get perfect scores) or extremely difficult (so difficult that most of the scores are zero). See also the discussion on "Skewed Distributions."

An entirely different configuration would emerge if we were to measure the standing height of humans, because there are two physical types of humans, male

* On the other hand, our decimal numbering system makes an interval size of 10 the most convenient, especially during the tallying process. For that reason, an interval of 10 or some multiple thereof is worth considering (even though it does not give you midpoints that are whole numbers) if it yields an appropriate number of class intervals (somewhere between 10 and 20) and if you are faced with so many scores that convenience in tallying is an issue.



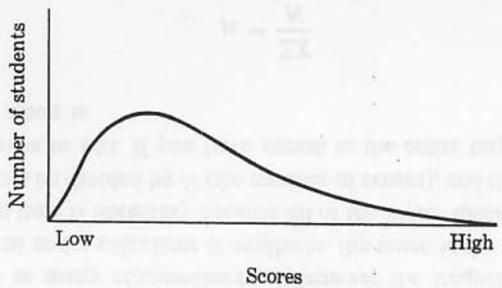
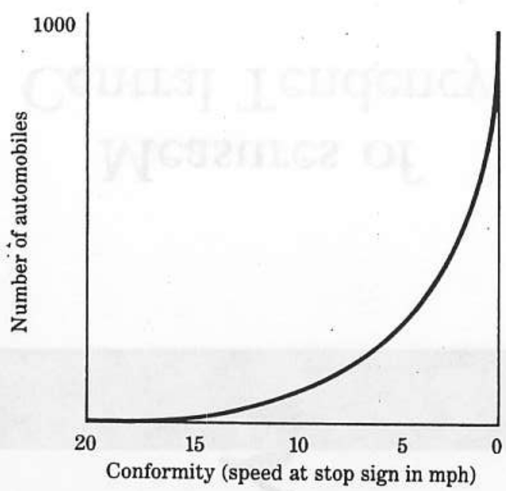


FIGURE 2-10 Positively skewed distribution.

and female. Thus, we should obtain a *bimodal* distribution, as shown in Figure 2-12. Similarly, if your instructor were to spring a pop quiz on Chapter 9 of this book at a time when only half of the class had read it, the distribution of those scores would be bimodal. (Note that in Figure 2-12 the distribution is apparently quite symmetrical; however, a bimodal distribution can be asymmetrical, as indeed this one surely would be if, say, it were composed of twice as many women as men.)

The J distribution is difficult to deal with statistically, and a bimodal distribution can be dealt with by separating the two normal distributions that are partially concealed within it. So these other configurations are of only passing interest to us here.



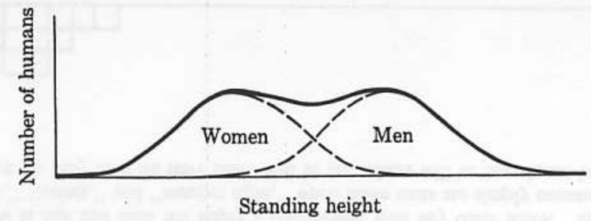


FIGURE 2-12 Bimodal distribution of humans on a scale of height.

SUMMARY

Many frequency distributions in the social sciences are approximately normal; that is, in the typical distribution there are a few very low scores and a few very high ones, but the great mass of individuals tend to pile up on the middle scores. This occurs because the probability of all the many determiners of a trait pointing in the same direction is virtually nil, and that of a balanced combination of determiners is much higher. Although other configurations do occur (positive skew, negative skew, J curve, and bimodality are mentioned), our primary concern in this book will be the normal distribution, because it serves as a good approximation to the kinds of distribution most frequently encountered in behavioral and medical investigations. Moreover, it is the configuration for which the best-known statistical treatments are available. We shall encounter some of those treatments in subsequent chapters.

There are many possible ways of organizing and displaying data. One is to "group" frequencies—to pool individual observations into class intervals. Grouping has both positive and negative consequences. A negative consequence is the loss of information that occurs when scores of varying values are assigned whatever value is at the midpoint of a given class interval. A positive consequence is that grouping tends to reveal underlying patterns by allowing many random variations to cancel each other. If you have a large number of observations to organize, another positive effect is that grouping can simplify calculations.

3

Measures of Central Tendency

Somewhere in the intermediate grades you were introduced to the concept of *average*, and you may have used it ever since on the assumption that there is only one such. Actually, there are several types of averages, of which three of the most common will be described here: the mean, the median, and the mode.

THE MEAN (μ AND $\overline{X}$)

The average that you learned about in grade school was a *mean*. The mean is not, as you were led to believe, *the* average, but it does have characteristics that make it the best one to use in many circumstances. Whenever the frequency distribution is fairly symmetrical and a calculator is available, the mean is the statistic of choice. The computation time is necessary because all of the scores must be added together before the sum can be divided by N (the number of scores), and there are sometimes thousands of scores to add. If you have access to the entire target population, the formula for the mean is

$$\mu = \frac{\Sigma X}{N} \qquad (3\text{-}1)$$

and X refers to the *raw scores* that have been recorded.* The expression ΣX (read "summation ecks"), therefore, is the sum of all the raw scores in the distribution. N is the size of the population.

If you do *not* have access to the entire population, the scores that you *do* have constitute a sample, and the formula for *its* mean is

$$\overline{X} = \frac{\Sigma X}{n} \qquad (3\text{-}2)$$

where $\overline{X}$ is the mean of a sample and n is the size of that sample. Notice that *the operations specified by the two formulas are exactly the same.* If you calculate what you believe to be a population mean using Formula (3-1) and subsequently learn that the scores you have comprise only a small fraction of the target population, all you need to do is to change the label from μ to $\overline{X}$; no further calculation is necessary.

Because in published research the description of samples is more common than that of populations, every reference to "the mean" after this chapter will be to a sample mean unless otherwise noted (although your own studies may frequently be of populations).

In any event, Formulas (3-1) and (3-2) say the same thing, and what they say has two aspects. The first is that the formulas *define* the mean; the second is that they specify a procedure for *calculating* it. But this book is not about calculating; our primary objective is rather to acquire an understanding of statistical concepts. The concept of the mean can best be understood through the following illustration.

Essence of the Concept

Try to imagine a long beam made of an exotic metal that is totally rigid and utterly weightless. Upon that beam we shall place 14 cubes (to represent 14 scores), each of equal weight. Figure 3-1 shows one possible distribution of cubes.

At what point on the beam would a fulcrum (support) have to be placed to establish an equilibrium? That is, what is the *balance point* of the distribution? Because this distribution is symmetrical, it should be easy to see that it would

* A *raw score* is one that does not imply a comparison with any other score; "inches," "points," "runs," "hits," "errors," and "number right," when those units are simply counted, are raw scores. Raw scores are the only kind we have dealt with so far, others will be introduced later.

balance if the fulcrum were placed at 5.5. But that is what all the computation is about; by using the formula, we can find the precise point we are seeking* even in cases where a diagram would be rather puzzling. (That such situations do arise will become clear in the section on "The Median.")

Another way of saying "balance point" is "the point from which all deviations sum to zero." In Figure 3-1, for example, that point is 5.5. One of the scores (the 2) deviates $3\frac{1}{2}$ units in the negative direction (left) and one (the 9) deviates $3\frac{1}{2}$ units in the positive direction (right); one is a negative and one is a positive $2\frac{1}{2}$; there are two negative and two positive deviations of $1\frac{1}{2}$; finally there are three scores $\frac{1}{2}$ unit below and three scores $\frac{1}{2}$ unit above the mean. Most distributions will not be perfectly symmetrical as this one is, but all will produce the same result: The positive and negative deviations from the mean will always cancel each other, bringing their sum to zero.

What I have been saying in different ways is that *the mean is the balance point* of any distribution of scores.

Appropriate Applications

As the balance point of a distribution, the mean is the only measure of central tendency that is sensitive to all of its scores. Probably its most important application is to other measures. Because of that balance-point feature, it is compatible with many more complex measures that you will meet later. Indeed, calculating a mean is an integral part of calculating a *standard deviation*, a product-moment *coefficient of correlation*, and all of the various *standard errors*, to name a few.

A related advantage of the mean over other measures of central tendency is its usefulness in making inferences from sample to population: The mean of a sample is the best estimate of that of the population. But even the best estimate will probably miss the mark, and is is important to know the probable extent of the error. The mean lends itself to error estimation as well, as you will see in Chapter 8.

So if you have a sample, and you want to know the population mean, you will choose the mean to represent central tendency in your sample. And if you anticipate calculating more complicated statistical measures, you will probably need to calculate your sample mean first.

As I mentioned in the preface, some people do not feel comfortable with a quantitative concept until they have followed the relevant computation. If you are one of those people, Box 3-1 provides you with an opportunity to do that. The calculation is based on the defining equation for the mean. Calculational formulas —and hence calculations—frequently differ from definitional ones (see pages 5–6), and when they do they tend to obscure the *meaning* of the operations they

0026-15

Box 3-1 Calculation of a Mean (see Figure 2-1)

(1) Class interval	(2) Midpoint X_m	(3) f	(4) fX_m
65–69	67	1	67
60–64	62	2	124
55–59	57	4	228
50–54	52	13	676
45–49	47	24	1128
40–44	42	36	1512
35–39	37	40	1480
30–34	32	36	1152
25–29	27	24	648
20–24	22	13	286
15–19	17	4	68
10–14	12	2	24
5–9	7	1	7
		$\Sigma = 200$	$\Sigma = 7400$

Column 1: Class intervals. See Figure 2-1, page 13.

Column 2: Midpoints of class intervals. When data are grouped into class intervals, all individuals (X) within each interval are treated as though they were at the midpoint of that interval (X_m). Of course, that is not strictly true, but if the class intervals are small, error is negligible.

Column 3: Frequency (f) of scores in each class interval.

Column 4: Product of midpoint and frequency (fX_m). Only one person scored in the 5–9 interval, and the midpoint of that interval is 7, so the fX_m for that interval is just 7. There are two scores in the 10–14 interval and its midpoint is 12, so its fX_m is $2 \times 12 = 24$. Four persons scored somewhere in the interval 15–19, so fX_m is $17 \times 4 = 68$, and so on through the remaining intervals.

$$\bar{X} = \frac{\Sigma X}{n}$$

It should be clear that the sum of column 3 is n. It should also be clear that simply adding all the individual X_m's will give us the same sum as that of column 4; that is, the sum of column 4 is the ΣX in the formula, give or take an error of negligible magnitude.

$$\bar{X} = \frac{7400}{200} = 37$$

represent. Since meaning is your mission here, it will be better for you to keep your calculations close to your concepts.

THE MEDIAN (Mdn)

Another way of indicating central tendency is to tell which point on the baseline divides the distribution into two equal parts. Note that I said it divides the *distribution* in half, not the *baseline*. Look back at Figure 3-1. There, the beam is the baseline; it is 35 units long, but $\frac{35}{2} = 17.5$ is not the median. Nor is the median defined as a point halfway between the lowest point (1.5) and the highest (9.5), although because of the perfect symmetry of the distribution, it happens to be there in that particular case.

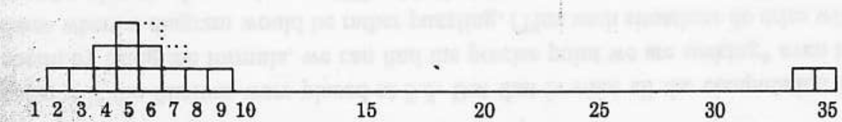
Essence of the Concept

In every case, the *median* is the point between the lower and upper halves of the distribution. (Remember, the distribution is the group of individual scores comprising the sample.) In Figure 3-1, that point is 5.5, because there are 7 scores below it and 7 above. In *any* distribution with an N of 14, the median will be midway between the 7th and 8th scores (counting up from the bottom of the distribution); in any distribution of 1000 individuals, the median will be the point midway between the 500th and 501st; and so on, with as many illustrations as you care to cite.[1]

Appropriate Applications

"Who is in what half of the distribution?" If that is our question, the answer will depend upon our first finding the median. A more important characteristic of the median, however, is that although it is not sensitive to the exact location of every score in the distribution, it can be used in situations where the mean would be inappropriate.

It may not be too much of an oversimplification to say that the median should be used in preference to the mean whenever the shape of distribution departs radically from perfect *symmetry*. Consider the situation depicted in Figure 3-1. There, because the distribution is perfectly symmetrical, the mean and the median are at precisely the same place (5.5). Now look at Figure 3-2 to see what happens when that symmetry is disturbed. In that figure we have exactly the same distribution as in Figure 3-1 except that two of the scores have been shifted far to the right.



The balance point—that is, the mean—has shifted also; it is now 9.5, which, lying as it does above 12 of the 14 scores, is probably an inappropriate index of the central tendency of this distribution.

But what has happened to the *median* as a result of that shift of two scores? It hasn't moved at all! (Count the scores above and below it, and see for yourself.) You may say that it *should* have moved—at least a little bit—because the distribution is different from what it was before; however, even though it is insensitive to that change, I'm sure you will agree that in this distribution the median is a better measure of central tendency than the mean, because the two extreme scores have *too much* influence on the mean. Often what has happened in such cases is that scores have been forced into one category when they should have been classified into two or more. For example, if you were to plot a distribution of the annual incomes of a football coaching staff, you would probably find that most are rather close together but that the salary of the head coach is distinctly separate from the others. If you had to report a single average of the salaries of football coaches at Central University, would you use the mean or the median? If you *don't* have to confine your report to a single central tendency, it probably would be better to report the head coach separately in this case, but if salaries of the senior members of the support staff approach that of the head coach, the median of the entire staff might be an appropriate index of central tendency.

In another situation that militates against using the mean, the scale is not long enough to accommodate all of the scores at one end of the distribution. For example, whereas a long, difficult test might produce a normal distribution of student scores, a short, easy one might pile up, say, a third of the scores at the top of the scale. The true magnitude of these scores is indeterminate: There is no way to know "how far out on the bar" each weight should be placed and hence no way to locate the balance point of the distribution. But you can use the median.

Finally, there is the rather rare array of categories that are essentially nonquantitative but nevertheless appear in a universally recognized *order*. Military ranks come to mind, but any ranking structure could illustrate the genre. An especially good one, though, is the results of a race—say a marathon—in which results are recorded in two forms: (1) the precise *time* that it took each contestant to reach the finish line and (2) the *order* in which they all reached it. If you know the runners' *times*, you can calculate their mean time, but if all you have is their *ranks*, the median will have to do, because there are no *scores* to support the calculation of a mean.

THE MODE

The last of the three averages to be presented here is also the easiest. The *mode* is simply the point with the greatest frequency. In Figure 2-1, the mode is 37; in

If your data are quantitative and ordered, as are nearly all of the data presented in this book, the mode's very simplicity explains one of its two main purposes: It is used when a very quick estimate is needed. It is also used specifically for identifying the typical (most common) score.

If your data are qualitative and not ordered, like sales of various colors of designer dresses or enrollments in the subject-matter categories of a college curriculum, the mode is really the only central tendency that *can* be used: You cannot count up from the bottom as you must in order to find a median, much less add scores as you must to find a mean. (There *are* no scores to add.)

The mode can therefore be used where the mean and the median cannot. Beyond that, it can supplement but not supplant either or both of those statistics.

SUMMARY

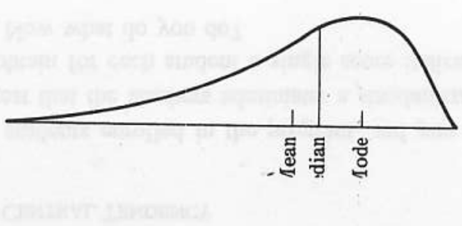
The three most common measures of central tendency are *mean*, *median*, and *mode*.

The *mean* is the balance point of a distribution. It is the only average that utilizes all available information, and when distributions are approximately normal it is the one that serves as an essential component of many of the more complex measures that you will encounter in later chapters.

The point at which a distribution can be cut in half is its *median*. It is used when precise location of the two halves is a prime concern, when data are ordered but not quantitative, or when a distribution of quantitative data is far from normal. In those circumstances the median can supplement or even supplant the mean.

The *mode* is the place where the greatest number of cases occurs; if the data are quantitative, it is the most common score. When that is exactly the information you want or when your data are neither quantitative nor ordered, the mode is the appropriate index of central tendency. In other situations it can be useful as a supplement to the median and/or the mean, but it should never supplant them.

In a skewed distribution, the arrangement (order) of the three measures along the baseline is predictable. If the skew is negative, as in Figure 3-3, the arrangement from left to right is: mean, median, and mode. If the skew is positive, as in Figure 3-4, the order is just the opposite: mode, median, and mean. Conversely, if you know the order of the three averages, you can tell the direction of the skew.



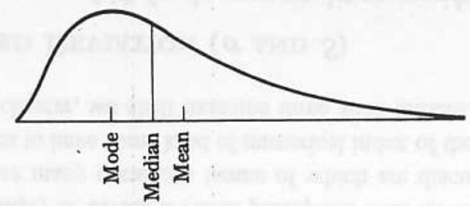


FIGURE 3-4 Measure of central tendency in a positively skewed distribution.

Because of the mean's affinity to many of the more complicated constructs that I will present later (beginning in Chapter 4), future references to measures of central tendency will be almost exclusively to the mean. For simply describing a small population, however, or for doing a preliminary exploration of data that might later be used to make inferences beyond your sample, all three measures can be helpful, as can graphic representations.

Sample Applications

The following problems require that you make some decisions about the appropriate use of statistics. In each case, you are presented with a situation that might be faced by a practitioner of the indicated discipline. Imagine yourself as that person in that situation.

Make a tentative choice without looking back into the chapter; explain it as best you can. Then review the text and refine your answer. Finally, check the back of the book (pages 165–181) for a suggested choice and brief discussion of it.

You may use that choice and that discussion to correct or further refine your response. But if you feel that yours is as good as mine, discuss it with your instructor or someone else who has a knowledge of statistics. If he or she agrees, I'd very much like to see what you have done.

Besides answering the questions in the following paragraphs, try to identify the sample and/or population that can be found in each.

EDUCATION

A cooperative vocational education program has recently opened to provide one year of training for 200 twelfth-grade students from 10 schools. The teachers are in reading skills of the students enrolled in the program, and you are called in as consultant. You suggest that the teachers administer a standardized reading comprehension test and obtain for each student a single score indicating the reading level of that student. Now what do you do?

POLITICAL SCIENCE

You are studying the domestic and military expenditures of European nations. You want to find the average amount spent by European countries on arms. You cannot afford to study all European countries, so you take a random sample. Now what do you do?

PSYCHOLOGY

You are a family counselor working with the mother of a three-day-old infant. The mother is very concerned about her child (her sister has recently given birth to an infant with birth defects) and asks you if her infant is showing normal behaviors for a newborn. You observe the infant in question, but you are not certain which behaviors are classified as normal for a neonate. Your task, therefore, is to find out how newborns tend to behave. You go to three hospitals in your city, visit the neonatal units, and observe and measure a variety of infant behaviors. For example, you dangle a large red ring in front of each infant to see whether it follows the ring visually or even attempts to grasp it. You sound a bell close to each infant's right ear and observe whether the infant turns toward the sound. You then assign points based on your observations (e.g., 1 point for following the ring visually and 2 points for grasping the ring). What single statistic would best represent all the children you have tested?

SOCIAL WORK

The director of a child welfare agency is interested in the length of time that families receive protective services. She asks you to provide information regarding the number of treatment hours received by these clients. To approach this problem you select a random sample of clients from the closed-case files. How can these data be analyzed?

SOCIOLOGY

A city council of a city of 100,000 wants to know the average income of its residents. You are asked to make an estimate but are given only a small budget for doing so. You draw a random sample of 100 residents and ascertain the average income of all residents. What computation is appropriate?

0026-18

4

Measures of Variability

Different populations (and the samples extracted from them) have different central tendencies, but they differ in another significant respect as well. Consider the two curves depicted in Figure 4-1. Both represent distributions of the same area (identical N's), and both have the same central tendency; nevertheless, the two distributions are very different. In what way are they different? You can see that one is spread out more than the other. Since the baseline on which the spreading occurs is a single scale of scores, the spreading means that the scores in that distribution *vary* more than those in the "squeezed together" distribution.

Diagrams provide a superior way of approaching a new concept, but even if we had a diagram of every distribution drawn to scale so that we could compare variabilities by inspection, there would still be a need (demonstrated in Chapter 5) for an index that can enter into mathematical operations that are essentially numerical; you can't multiply or divide a visual perception with an acceptable degree of precision. There are many occasions (some of which are discussed in Chapter 5) when it is important to have some kind of numerical index of the variability of a set of scores. In this chapter, we shall examine three such indices.

THE STANDARD DEVIATION (σ AND S)

Taking its importance on faith for the moment, let us consider how an index of

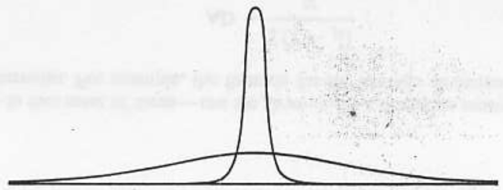
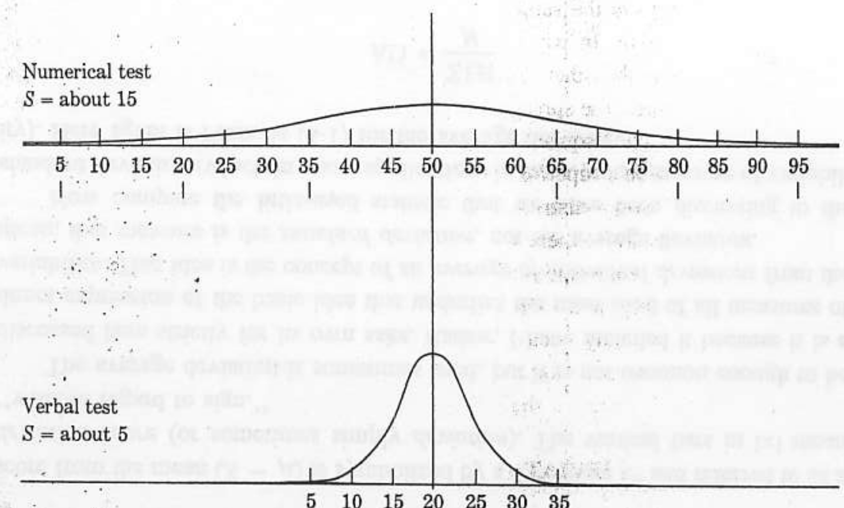


FIGURE 4-1 Two distributions with the same N's but different variabilities.

the single scholastic aptitude test mentioned in Chapter 2, those students took a *pair* of tests—one of verbal aptitude, the other of numerical. The distributions of scores on the two tests might look something like Figure 4-2.* In that diagram, the means of the two distributions have been aligned so that we may concentrate on their

* The indicated central tendencies and variabilities of the two distributions make somewhat plausible the proposition that if combined, they would form the distribution shown in Figure 2-1. However, that was not the primary consideration in their selection. Rather it was *simplicity*. We are deemphasizing computation; therefore, computations that *are* required have been made as easy as possible. It is easier to think about an interval that extends from 20 to 25, for example, than one that extends from 23 to 27, or even from 23 to 28. You will find that you can do most—possibly all—of this book's computations entirely in your head.

But do not attempt any computations unless the requisite data are readily available. For example, in Figure 4-2 do not attempt to confirm the standard deviations of 15 and 5 in the two distributions. To do so, you would need a list of the individual scores, and those have been withheld deliberately to encourage you to focus on the big picture—the configuration of the entire sample for each test and the comparison of two configurations that are very different from each other.



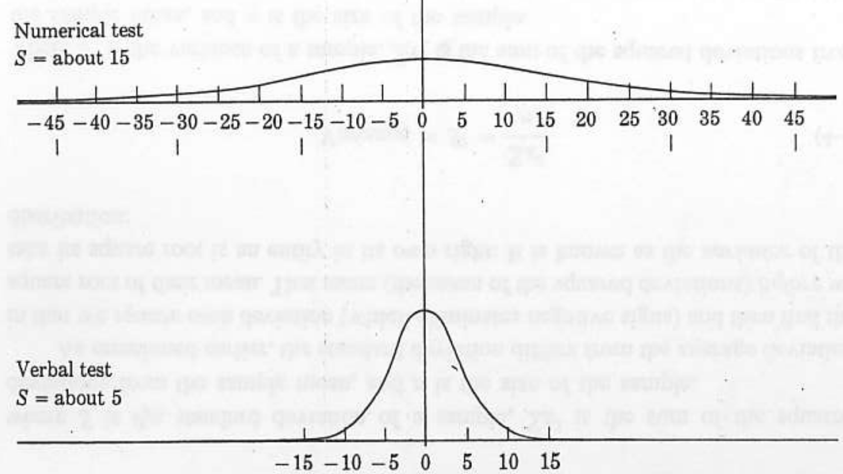
respective variabilities. As we look at the figure, we are immediately impressed by the striking difference between the dispersions of the two distributions across their baselines. But it is not enough to be impressed; we need a numerical index of dispersion (variability).

Essence of the Concept

How might such an index be devised? One way would be to compare every score in the sample to every other and to average the differences obtained thereby. But that would be entirely too cumbersome, especially with large distributions; we can get the same effect by selecting a point of reference in the middle of the distribution and measuring the distance of each individual score from that point, as in Figure 4-3. The average (mean) of those differences (without regard to sign) can also serve as an index of variability. If, for example, we were to choose the mean of our target population as our reference point, our index would be an average of the individual distances from the mean and could be obtained via the following formula:

$$\mathrm{AD} = \frac{\Sigma|x|}{N} \qquad (4\text{-}1)$$

where AD is the *average deviation*, $\Sigma|x|$ is the sum of *individual deviations* from the mean of the population, and N is the size of the population. The distance of any



score from the mean $(X - \mu)$ is symbolized by a lowercase x* and referred to as a *deviation score* (or sometimes simply *deviation*). The vertical bars in $|x|$ mean "without regard to sign."

The average deviation is sometimes used, but it is not common enough to be discussed here strictly for its own sake. Rather, I have included it because it is a direct expression of the basic idea that underlies the most used of all measures of variability. That idea is the concept of an *average of individual deviations* from the mean; that measure is the *standard deviation*, not the average deviation.

Now compare the little-used statistic that we have been discussing to the standard deviation (which in many applications is *the* accepted measure of variability). Here again is Formula (4-1) for the average deviation:

$$\text{AD} = \frac{\Sigma|x|}{N}$$

and here is the formula for the standard deviation:

$$\sigma = \sqrt{\frac{\Sigma x^2_{\text{pop.}}}{N}} \qquad (4\text{-}2)$$

where σ is the standard deviation of the population, Σx^2 is the sum of the deviations from the population mean, and N is the size of the population. The two formulas are very similar, are they not? In fact, they are exactly alike except that for the standard deviation we take a mean of *squared* deviation scores; then we take the *square root* of that mean. But don't worry about the differences between Formulas (4-1) and (4-2). The important thing is the similarity: The standard deviation is a kind of average of individual deviations from the mean of the distribution.

In Chapter 3 you learned that the defining equation for a sample mean is essentially the same as the one that defines the population mean. The same is true of sample and population standard deviations, as you will see immediately when you compare Formula (4-2) with this one:

$$S = \sqrt{\frac{\Sigma x^2_{\text{sample}}}{n}} \qquad (4\text{-}3)$$

* Some authors—in fact most of them—use the *formula* for a deviation score as the *symbol* for it that appears in other formulas. For example, the formula for the average deviation would be

$$\text{AD} = \frac{\Sigma|X - \mu|}{N}$$

where S is the standard deviation of a sample, Σx^2 is the sum of the squared deviations from the sample mean, and n is the size of the sample.

As mentioned earlier, the standard deviation differs from the average deviation in that we *square* each deviation (which eliminates negative signs) and then find the square root of their mean. That mean (the mean of the squared deviations) *before* we take its square root is an entity in its own right: It is known as the *variance* of the distribution:

$$\text{Variance} = S^2 = \frac{\Sigma x^2}{n} \qquad (4\text{-}4)$$

where S^2 is the variance of a sample, Σx^2 is the sum of the squared deviations from the sample mean, and n is the size of the sample.

Let us take a moment to sum up: In this and the previous chapter we have examined a sequence of concepts that share a common structure. It is important for you to be aware of that structure; not only because it will help you to understand its manifestations in concepts developed earlier but also because it will appear in other concepts later. The best way to discern the structure is to place all the items of the series in close proximity to one another. Here are the four concepts in order in which they were presented:

$$\text{Mean of the raw scores} = \frac{\Sigma X}{n}$$

$$\text{Average deviation} = \frac{\Sigma|x|}{n}$$

$$\text{Variance} = \frac{\Sigma x^2}{n}$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma x^2}{n}}$$

Their common structure should be apparent by inspection. They are all *averages*. The first is an average of raw scores; the other three are averages of deviation scores. In every case, the average is a mean. In the case of variance it is the mean of the *squared* deviation scores. Last is the standard deviation, which is the *square root* of the variance and therefore is expressed in linear (*not* squared) units.

Appropriate Applications

The standard deviation is to measures of variability what the mean is to measures of central tendency. It is sensitive to the value of every score, and in a normal both of which will be discussed later. In scientific work, at least, the standard deviation clearly is the prevalent measure of dispersion.

As in Chapter 3 a calculation box (Box 4-1) provides you with an opportunity to try your hand at actually crunching some numbers. But here again, and for the same reason as before, the operations in the box are specified by the defining equation, not by a specialized calculating formula.

The Interquartile Range (IQR)

A much easier statistic to comprehend (and to compute) is the *interquartile range* (IQR). Look at Figure 4-4. What you see is a negatively skewed distribution cut into four equal parts. At the upper end of each of those quarters is a point on the baseline called a *quartile* (Q)—the first quartile (Q_1) above the lowest quarter, the second (Q_2) above the lowest two quarters,* and the third (Q_3) above the lowest three quarters. The point just above the highest score in the distribution would logically be Q_4, but that expression is seldom if ever used.

Essence of the Concept

The *interquartile range*, like any measure of variability, is an interval. In this case the interval extends from Q_1 to Q_3. The calculation of the interquartile range is extremely easy: You simply find the difference between Q_1 and Q_3. That's it.[1]

The IQR, like the median, is insensitive to the precise values of most of the scores in a distribution, so when you use it instead of the standard deviation you discard information. But if the distribution is skewed, you can return some important information (namely, the direction of the skew) by reporting not only the *difference* between Q_1 and Q_3, but their exact *locations* as well, along with that of Q_2. (*If* $Q_2 - Q_1$ is longer than $Q_3 - Q_2$, the skew is negative, as in Figure 4-4; if $Q_3 - Q_2$ is the longer, the skew is positive.)

Appropriate Applications

The interquartile range is to measures of variability what the median is to measures of central tendency. Although insensitive to the exact values of many of the scores in a distribution, it is preferred to the standard deviation in the same kinds of situations in which the median is preferred to the mean—namely, when distributions are radically asymmetrical. It is the perfect companion to the median wherever the latter is properly applied.

A report of the incomes of a university faculty, for example, might be skewed by the high salaries and consulting fees of the college of business. If faculty

* What other familiar statistic has essentially the same definition as the second quartile? If you are not

0026-21

Box 4-1 Calculation of Standard Deviation (see Figure 4-2, verbal test)

(1) Class interval	(2) Midpoint X_m	(3) Frequency f	(4) $X_m - \overline{X}$ x	(5) $(X_m - \overline{X})^2$ x^2	(6) $f(X_m - \overline{X})^2$ fx^2
33–37	35	2	15	225	450
28–32	30	13	10	100	1300
23–27	25	32	5	25	800
18–22	20	106	0	0	0
13–17	15	32	−5	25	800
8–12	10	13	−10	100	1300
3–7	5	2	−15	225	450
		$\Sigma = 200$			$\Sigma = 5100$

Column 1: Class intervals. See pages 12 and 16–21.

Column 2: Midpoints of class intervals. Remember when data are grouped into class intervals, all individuals (X) within each interval are treated as though they were at the midpoint of that interval (X_m).

Column 3: Frequency (f) of scores in each class interval.

Column 4: Deviation scores (x), which equal the differences between the midpoints (X_m) and the mean of the distribution ($\overline{X}$). In this case, $\overline{X} = 20$.

Column 5: Squares (x^2) of the deviation scores listed in column 4.

Column 6: Product of the squared deviation scores and the frequency of scores in the interval (fx^2).

$$s = \sqrt{\frac{\Sigma x^2}{n}}$$

The sum of the frequencies listed in column 3 is n, and the sum of column 6 is Σx^2.

$$s = \sqrt{\frac{5100}{200}} = 5.01$$

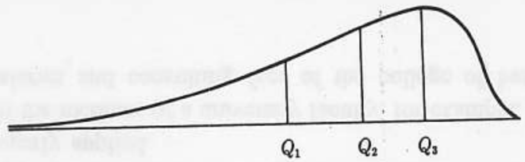


FIGURE 4-4 Negatively skewed distribution with area divided into quarters.

incomes without the college of business are distributed symmetrically, the high incomes there give the total distribution a marked positive skew, making the mean and standard deviation difficult to interpret. The median and interquartile range would be better in those circumstances than the mean and standard deviation.

THE RANGE

There is another measure of variability—one that probably gets more attention than it deserves, both in makeshift analyses and in this book. It is given attention in makeshift analyses because it is so easy to compute, and much of the space assigned to it in this book is occupied by an exposure of its fundamental weakness and an attempt to prepare you for possible differences in its interpretation.

Essence of the Concept

Sometimes we are interested specifically in the most extreme cases in a sample; on these occasions we may report its *total range*—or simply *range*—which is the distance from the lowest score to the highest. As you can see, it is extremely easy to calculate.[2] That is about its only virtue, however.

In fact, the most important feature of the range is a weakness—its extreme instability. Note that the range in Figure 3-1 is 8. Now turn to Figure 3-2. It is the same as Figure 3-1 with the exception that two scores have been moved away from the main group. Note the effect on the range. (Instead of 8, it is now 35.5 − 1.5, or 34!) Note, too, that not even *two* extremely high scores were necessary to have that effect; one would have done exactly the same thing. The fact is that the range is determined by two, and only two, scores in any distribution: the lowest and the highest. (Compare that to the standard deviation, which is sensitive to *all* scores.) That is why the range is so easy to compute. That is also why it is so unreliable.

Appropriate Applications

We have seen that the standard deviation is analogous to the mean and is its proper companion, and we have noted a similar relationship of the interquartile range to the

One similarity is that each is the *quickest* estimate of its kind. Another closely related similarity is that each is *less stable* than alternative indices (although the range usually is much the worse in that respect because of its total dependence on only two scores). Finally, both the mode and the range are used, more often than are other statistics, to answer questions related directly and very simply to their definitions. For the mode, the question is "Which is the typical (most frequent) case?" For the range, it is "How much of the scale must be used to represent the distribution?"

SUMMARY

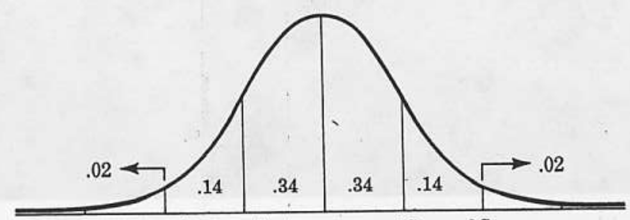
Two important ways of describing a sample are by its central tendency or average, which indicates the general level of the scores, and by its variability, which tells the extent to which individual scores deviate from that average. This chapter has been about the latter type of description—measures of *variability*.

You have been invited to use your previously developed understanding of measures of central tendency as an anchor for the new concepts; the latter were presented as analogs of the former. Specifically, the standard deviation is roughly analogous to the mean and is its companion statistic, the interquartile range goes with the median, and the range is similar to the mode.

The *standard deviation* is a kind of average of individual deviations from the mean of a distribution. The mean of the squared deviations is called the *variance*, and the standard deviation is the square root of the variance. In a normal distribution, the standard deviation carries the most information of all measures of variability. It also lends itself to the computation of many more advanced measures.

The *interquartile range* is the distance from Q_1 to Q_3. It embodies less information than the standard deviation but is preferred to it whenever the distribution is markedly asymmetrical, and it is the perfect companion to the median.

The *range* is the distance from the lowest score to the highest, is completely determined by those two scores, and thus carries little information. Like the mode, it



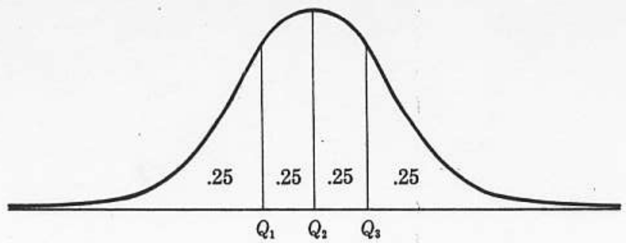


FIGURE 4-6 Normal distribution with area divided into quarters.

is both easier to compute and less stable than any other measure of its kind. It is, of course, the one statistic that gives information specifically about the distance between the highest and lowest scores in the sample.

We have seen that in a normal distribution, the mean, the median, and the mode are all at the same place. When we make a similar graphical comparison of measures of variability, the situation is not quite as simple.

In Figure 4-5, the baseline is divided into equal units—namely, standard deviations—and the areas subtended by them (.02, .14, .34, etc.) are *un*equal. The number above the line indicates the proportion of the total area that is subtended by each segment. The proportions have been rounded off because they are easier to remember in that form, and they are precise enough for our present purposes in any case (more exact proportions are given in the figure in note 3 of Chapter 5).

In Figure 4-6, the baseline is divided into unequal segments, and it is the parts of the area that are equal. And, of course, the range subtends the entire distribution, as shown in Figure 4-7.

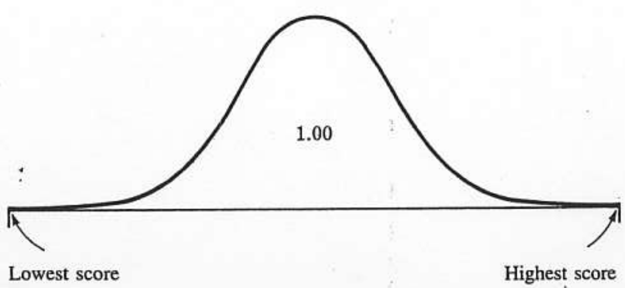


FIGURE 4-7 Normal distribution showing total range on baseline.

Sample Applications

EDUCATION

You are appointed to a committee of elementary school teachers and administrators. The committee decides to make the teaching of reading a top priority for the next two-year period. You get permission to use the annual in-service training fund for purchasing a reading program in which all elementary teachers will participate. After an extensive search, you narrow the field to two programs that have been used and evaluated on usefulness and practicality by a large number of elementary school teachers. The means of their ratings of the two programs are approximately the same. Is there any other statistic that might help you make your decision?

POLITICAL SCIENCE

You are interested in the incidence of military coups in Latin America. Specifically, you want to know whether most Latin American countries have experienced a number of coups that is close to the mean number for the region. After you have gathered your data, how do you obtain that information?

PSYCHOLOGY

You are one of a five-person team of observers sent into a home to evaluate the degree of aggressiveness among family members over a one-week period. The observers indicate that, on the average, eight aggressive acts occur per day. But there is also some disagreement. How might you quantify that disagreement?

SOCIAL WORK

You are the new director of a community fund-raising organization. The member agencies have widely varying needs, but you suspect that recently the board has been shirking its duty to investigate those needs. Specifically, you suspect that recent allocations have not been sufficiently differentiated. How might you document your case?

SOCIOLOGY

In order to make some decisions about the construction of family dwellings in your state, a construction firm asks you to report on the variability of family size. You have access to all of the state's data on family size. What measure of dispersion do