Chapter Outline

ORDINARILY, psychologists conduct research to test a theoretical principle or the effectiveness of some practical procedure. For example, a psychophysiologist might measure changes in heart rate from before to after solving a difficult problem. These measurements would then be used to test a theory that predicts that heart rate should change following successful problem solving. An applied social psychologist might examine the effectiveness of a program of neighborhood meetings intended to promote water conservation. Such studies are conducted with a particular group of research participants, but the researchers use inferential statistics to make more general conclusions about the theoretical principles or procedure being studied. These conclusions go beyond the particular group of research participants studied.

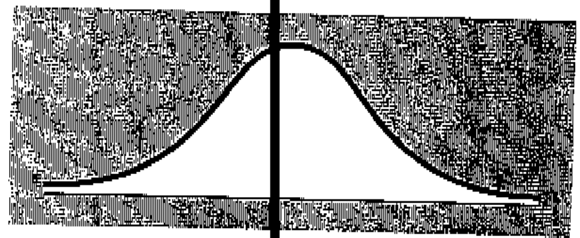
This chapter and Chapters 6, 7, and 8 introduce inferential statistics. They are the foundation for most of the rest of what you will learn in this book. In this chapter, we consider three topics: the normal curve, probability, and population versus sample. This is a comparatively short chapter, preparing the way for the next ones, which are more demanding.

THE NORMAL DISTRIBUTION

We noted in Chapter 1 that the graphs of many distributions of variables that psychologists study (as well as many other distributions in nature) follow a unimodal, roughly symmetrical, bell-shaped distribution. These bell-shaped histograms or frequency polygons approximate a precise and important mathematical distribution called the **normal distribution,** or more simply,

normal distribution

FIGURE 5-1
A normal curve.



normal curve

the **normal curve.**[1] (It is also often called a *Gaussian distribution* after the astronomer Karl Friedrich Gauss. However, if its discovery can be attributed to anyone, it should really be to Abraham De Moivre—see Box 5-1.) An example of the normal curve is shown in Figure 5-1.

Why the Normal Curve Is So Common in Nature

Take, for example, the number of random letters a particular person can remember accurately on various testings (with different random letters each time). On some testings the number of letters remembered may be high, on others low, and on most somewhere in between. That is, the number of random letters a person can recall on various testings probably approximately follows a normal curve. Suppose that the person has a basic underlying ability to recall, say, seven letters in this kind of memory task. Nevertheless, on any particular testing, the actual number recalled will be affected by various influences, such as noise in the room, the person's mood at the moment, a combination of random letters unwittingly confused with a familiar name, a sequence of random letters that happens to be mostly the same letter, and so on.

These various influences add up to make the person do better than seven on some testings and worse than seven on others. However, the particular combination of such influences that occur at any testing is probably essentially random. Therefore, on most testings positive and negative influences should roughly cancel out. The chances of all the negative influences happening to come together on a testing when none of the positive influences show up is not very good. Thus, in general, the person remembers a middle amount, an amount in which all the opposing influences cancel each other out. Very high or very low numbers of letters remembered are much less common.

This creates a distribution that is unimodal—most of the cases near the middle and fewer at the extremes. It also creates a distribution that is symmetrical, because a score is as likely to be above as below the middle. Being a

[1]The formula for the normal curve (when the mean is 0 and the standard deviation is 1) is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

where $f(x)$ is the height of the curve at point x and π and e are the usual mathematical constants (approximately 3.14 and 2.72, respectively). However, psychology researchers almost never use this formula because it is built into the various computer programs that do statistical calculations involving normal curves. And when work must be done by hand, any needed information about the normal curve is provided in tables in statistics books (for example, Table B-1 at the back of this book).

unimodal, symmetrical curve does not guarantee that it will be close to a normal curve; its tails could be too high or too low. However, it can be shown mathematically that in the long run, if the influences are truly random, a precise normal curve will result. (The proof can be found in a mathematical statistics text.) Mathematical statisticians call this principle the *central limit theorem*. We have more to say about this principle in Chapter 7.

The Normal Curve and Percentage of Cases Between the Mean and 1 and 2 Standard Deviations From the Mean

Because the shape of the normal curve is standard, there is a known percentage of scores below or above any particular point. For example, exactly 50% of the scores are below the mean. This is because in any symmetrical distribution, half the scores are below the mean. More interestingly, as shown in Figure 5-2, approximately 34% of the scores are always between the mean and 1 standard deviation from the mean. (Notice, incidentally, that in Figure 5-2 the 1 standard deviation point on the normal curve is at the place on the curve where it starts going more out than down.)

To illustrate the usefulness of the normal curve being completely standard, consider IQ scores. On many widely used intelligence tests, the mean IQ is 100, the standard deviation is 16, and the distribution of IQs is considered to be approximately normal (see Figure 5-3). Knowing about the normal curve and the percentage of scores between the mean and 1 standard deviation above the mean allows us to know that about 34% of people have IQs between 100, the mean IQ, and 116, the IQ score 1 standard deviation above the mean. Because the normal curve is symmetrical, about 34% of people have IQs between 100 and 84 (the score 1 standard deviation below the mean), and 68% (34% + 34%) have IQs between 84 and 116.

There is something else you can see by looking at the normal curve: There are many fewer scores between 1 and 2 standard deviations from the mean than there are between the mean and 1 standard deviation from the mean. About 14% of the scores are between 1 and 2 standard deviations above the mean (see Figure 5-2). Similarly, because the normal curve is symmetrical, about 14% of the scores are between 1 and 2 standard deviations below the mean. Thus, about 14% of people have IQs between 116 (1 standard deviation above the mean) and 132 (2 standard deviations above the mean).

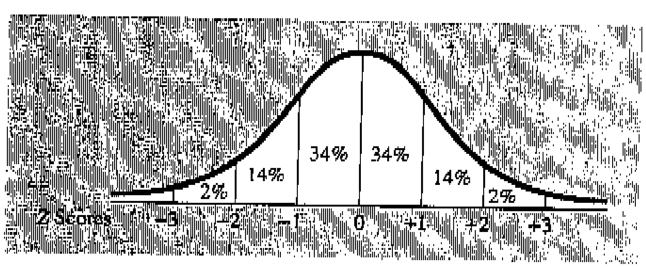


FIGURE 5-2
Normal curve with approximate percentages of scores between the mean and 1, 2, and 3 standard deviations above and below the mean.

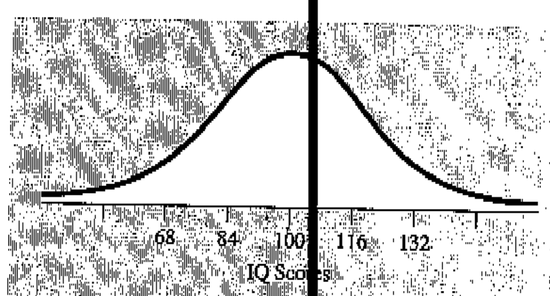


FIGURE 5-3
Distribution of IQ scores on many standard intelligence tests (with $M = 100$ and $SD = 16$).

You will find it very useful to remember the 34% and 14% figures. Thes[e] figures tell you the percentage of people above and below any particular scor[e] whenever you know that score's number of standard deviations above o[r] below the mean.

It is also possible to reverse this approach and figure out a person's number of standard deviations from the mean from a percentage. Suppose you are told that a person scored in the top 2% on a test. Assuming that scores on the test are approximately normally distributed, the person must have a score that is at least 2 standard deviations above the mean. This is because of the 50% of the scores above the mean, 34% are between the mean and 1 standard deviation above the mean and another 14% are between 1 and 2 standard deviations above the mean. That leaves 2% (that is, 50% – 34% – 14% = 2%).

Similarly, suppose you were selecting animals for a study and needed to consider their visual acuity. Suppose also that visual acuity was normally distributed and you wanted to use animals in the middle two thirds (a figure close to 68%) for visual acuity. In this situation, you would select animals that scored between 1 standard deviation above and 1 standard deviation below the mean. If you knew the mean and the standard deviation of the visual acuity test, you could then go on to pinpoint the raw score lowest and highest acuity levels.

The Normal Curve Table and *Z* scores

The 34% and 14% figures are useful practical approximation rules, even when a score is not exactly at 1 or 2 standard deviations from the mean. These percentages give you a general sense of where any particular score stands in relation to the other scores in the distribution. However, in many research and applied situations, psychologists need more precise information. Fortunately, because the normal curve is exactly defined, such precision is feasible. It is possible to compute, for example, the exact percentage of scores between any two points on the normal curve, not just those that happen to be right at 1 or 2 standard deviations from the mean. That is, it is possible to determine the exact percentage of scores between any two *Z* scores. For example, exactly 68.59% of scores have a *Z* score between +.62 and –1.68; exactly 2.81% of scores have a *Z* score between +.79 and +.89; and so forth.

It is possible to compute these exact percentages applying the formula for the normal curve and integrating, using calculus. However, in practice, psychologists do this much more simply. Statisticians have worked out tables for the normal curve that give the percentage of scores between the mean (a *Z* score of 0) and any other *Z* score. Suppose you want to know the percentage of scores between the mean and a *Z* score of .62. You simply look up .62 in the table and it tells you that 23.24% of the scores are between the mean and this *Z* score.

We have included such a **normal curve table** in Appendix B (Table B-1). As you can see, the table has two columns. The first column lists the *Z* score; the column next to it gives the percentage of scores between the mean and that *Z* score. Notice also that the table repeats these two columns several times on the page. Be sure to look across only one column. Also, notice that the table lists only positive *Z* scores. This is because the normal curve is perfectly symmetrical. Thus, the percentage of scores between the mean and,

normal curve table

say, a Z of +2.38 is exactly the same as the percentage of scores between the mean and a Z of −2.38.

In our example, you would find .62 in the "Z" column and then, right next to it in the "% mean to Z" column, you would find 23.24.

You can also reverse the process and find the Z score that goes with a particular percentages of scores. Suppose you were told that Janice's creativity score was in the top 10% of ninth grade students. Let's also assume that creativity scores follow a normal curve. You could figure out her Z score as follows: First you would reason that if she is in the top 10%, 40% of the students have scores between her score and the mean. (There are 50% above the mean and she is in the top 10% of scores overall. This leaves 40%.) Then, you would look at the "% mean to Z" column of the table until you found a percentage that was close to 40%. In this case, the closest you could come would be 39.97%. Finally, you would look at the "Z" column to the left of this percentage. The Z score for 39.97% is 1.28. If you know the mean and standard deviation for ninth-grade students' creativity scores, you could figure out Janice's actual raw score on the test. You would do this by changing her Z score of 1.28 to a raw score using the usual method of changing Z scores to raw scores.

Procedures for Finding the Percentages of Scores From Raw Scores and *Z* Scores Using a Normal Curve Table

Based on the above discussion, we can now systematically review the procedures for finding percentages of scores from Z scores. If you are working with raw scores, change them to Z scores, using the procedures described in Chapter 2. Then proceed as follows.

First draw a picture of the normal curve. Where the Z score falls on it, shade in the area for which you are trying to find the percentage. Then estimate the percentage in the shaded area based on the 50%–34%–14% practical rule. When drawing in the Z score, be sure to put it in the right place above or below the mean according to whether it is a positive or negative Z score. Drawing a picture of the problem and making a rough guess are important, because you will be much less likely to make mistakes in the more precise figuring.

Once you have your picture and your rough guess, you can go on to find the exact number. The main step is to look up the Z score in the "Z" column of Table B-1 and find the percentage in the "mean to Z" column next to it. If you want the percent of scores between the mean and this Z score, this would be your final answer. But often you will need to add 50% to this percentage. This is necessary when the Z score is positive and you want the total percent below this Z score; or when the Z score is negative and you want the total percent above this Z score. Other times you will have to subtract this percentage from 50%. This is necessary when the Z score is positive and you want the percent higher than it; or when the Z score is negative and you want the percent lower than it.

It isn't necessary to memorize such rules. It is much easier to make a picture for the problem and figure out whether the percentage you have from the table is correct as is, or if you need to add or subtract 50%.

Examples

Consider some examples using IQ scores. Suppose a person has an IQ of 125. What percentage of people have higher IQs? Before proceeding we need to convert this to a *Z* score. Assuming a mean of 100 and a standard deviation of 16, an IQ score of 125 is equal to a *Z* score of +1.56. Now that we have the *Z* score, the first step is to draw the picture. In Figure 5-4 we have shaded in the area above a *Z* score of 1.56. Then we want to approximate the percentage using the 50%–34%–14% rule. A *Z* score of 1 has 16% of the scores above it. (This is because it has 34% of the scores between it and the mean, and there are 50% of the scores above the mean. That leaves 16% above 1 standard deviation.) As we saw in an earlier example, a *Z* score of 2 has 2% of the scores above it. Thus, a *Z* score of 1.56 will have somewhere between 16% and 2% of the scores above it.

Having drawn the picture and estimated the percentage, we are now ready to figure it exactly. In the normal curve table, 1.56 in the "*Z*" column goes with 44.06 in the "% mean to *Z*" column. Thus, 44.06% of people have IQ scores between the mean IQ and an IQ of 125 (*Z* score of +1.56). In a normal curve, 50% of people are above the mean. Since 44.06% of the people above the mean are below this person's IQ, that leaves 5.94% (50% – 44.06%) above this person's score. This is the answer to our problem (and is shown in Figure 5-4). Notice that this percentage is within the range we estimated using the 50%–34%–14% approximation rule.

Now consider a person with an IQ of 95. What percent of people have IQs higher than this person? Following the usual procedure for changing a raw score to a *Z* score, an IQ of 95 is equal to a *Z* score of –.31. Figure 5-5 shows our picture for this situation; note that we have shaded in the area of the curve above a *Z* score of –.31. Our *Z* score is between 0 and –1. A *Z* score of 0 has 50% of the scores above it. A *Z* score of –1 has 84% of the scores

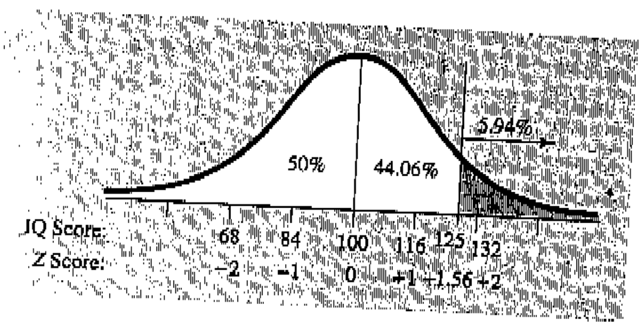


FIGURE 5-4
Distribution of IQ scores showing shaded region for scores above an IQ score of 125.

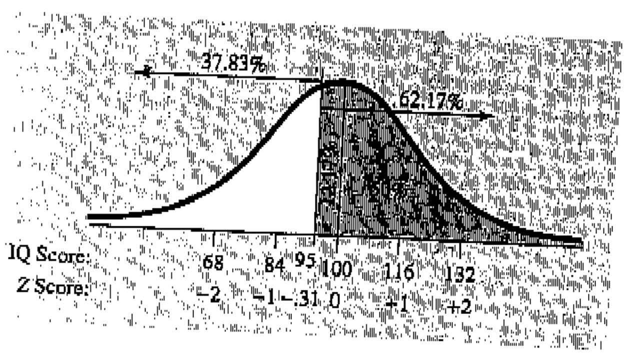


FIGURE 5-5
Distribution of IQ scores showing shaded region for scores above an IQ score of 95.

above it. (This is because there are 34% of the scores between −1 and 0, plus another 50% above 0, for a total of 84%.) Thus, our Z score of −.31 will have somewhere between 50% and 84% of the scores above it.

Now to the exact figuring. The normal curve table shows that 12.17% of scores are between the mean and a Z score of .31. Because the normal curve is symmetrical, this is also the area between a Z of −.31 and the mean. Thus, the total area above −.31 is 12.17% plus the 50% above the mean, for a total of 62.17%. (This is within our approximated range of 50% to 84%.)

Notice also that the percentage of scores *below* a Z score of −.31 would be the 50% below the mean *minus* the 12.17% between the mean and −.31. That leaves 37.81% of scores below a Z score of −.31.

Procedures for Finding Raw Scores and Z Scores From Percentages of Scores Using the Normal Curve Table

Going from a percentage to a Z score is similar to going from a Z score to a percentage. In both cases you begin by making a picture of the problem, shading in the approximate percentage, and making a rough guess of the Z score using the 50%–34%–14% figures. The rest of the process is almost exactly the reverse of going from a Z score to a percentage. Looking at the picture, you figure out the percentage between the mean and where the shading starts or ends. For example, if your percentage is the top 8% then the percentage from the mean to where that shading starts is 42%. If your percentage is the bottom 35%, then the percentage from the mean to where the shading starts is 15%. If your percentage is the top 83%, then the percentage from the mean to where the shading stops is 33%.

Once you have the percentage from the mean to where the shading starts or stops, you look up the closest number you can find to it in the "% Mean to Z" column of the normal curve table. The Z score in the "Z" column next to it will be your answer—except it may be negative. The best way to tell if it is positive or negative is from your approximation and by looking at your picture.

If you need a final answer in raw score terms, convert the Z score to a raw score using the methods you learned in Chapter 2.

Examples

Once again, we will use IQ scores for our examples. What IQ score would a person need to be in the top 5%? Figure 5-6 shows our picture; notice we have shaded the area with the top 5%. Using the 50%–34%–14% rule, we can figure that the Z score for the top 5% is between +1 and +2. We figure this as follows: Of the 50% above mean, 34% are between the mean and 1 standard deviation, leaving 16% above 1 standard deviation. However, because there are 14% between 1 and 2 standard deviations, there are only 2% left above 2 standard deviations.

Regarding the exact Z score, we first find the percent between the mean and where our shaded area starts. In this case, if 50% have IQs above the mean, at least 45% have IQs between this person and the mean (50% − 5% = 45%). Looking in the "% mean to Z" column of the normal curve table, the

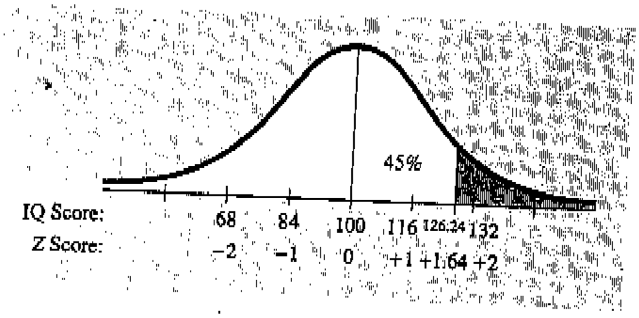


FIGURE 5-6
Z score and IQ raw score corresponding to the top 5%.

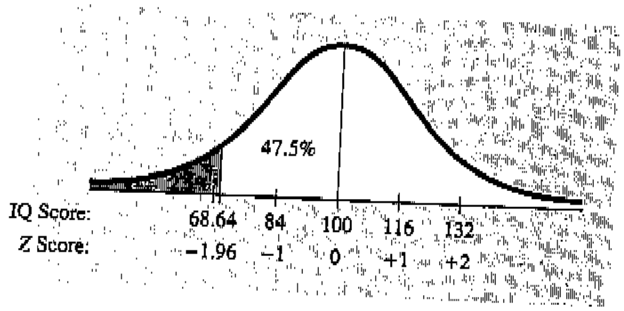


FIGURE 5-7
Z score and IQ raw score corresponding to the bottom 2.5%.

closest to 45% is 44.95% (or you could use 45.05%). This goes with a *Z* score of 1.64 in the "*Z*" column. As expected from our initial approximation, this answer is between +1 and +2.

To find the raw score, we can use the formula from Chapter 2: $X = M + (Z)(SD)$. With a mean IQ of 100 and a standard deviation of 16, you could conclude that to be in the top 5%, a person would need an IQ of at least 126.24.

Now consider what IQ would be in the lowest 2.5%. Our diagram of the problem, shading in the bottom 2.5%, is shown in Figure 5-7. The bottom 2% of a normal curve begins at 2 standard deviations below the mean (just as the top 2% begins at a +2). Thus, we can estimate that our answer will be somewhere near a *Z* score of −2. In more precise terms, the bottom 2.5% means that at least 47.5% of people have IQs between this IQ and the mean (50% − 2.5% = 47.5%). In the normal curve table, 47.5% in the "% mean to *Z*" column goes with a *Z* score of 1.96. Since we are below the mean, this becomes −1.96—a number quite close to our estimate of −2. Converting to a raw score, the IQ for the bottom 2.5% comes out to an IQ of 68.64.

PROBABILITY

The purpose of most psychological research is to examine the truth of a theory or the effectiveness of a procedure. But scientific research of any kind can only make that truth or effectiveness seem more or less likely; it cannot give us the luxury of knowing for certain. Probability is very important in science. In particular, probability is very important in inferential statistics, the methods psychologists use to go from results of research studies to conclusions about theories or applied procedures.

Probability has been studied for centuries by mathematicians and philosophers. Yet even today, the topic is full of controversy. Fortunately, however, you need to know only a few key ideas to understand and carry out the inferential statistical procedures you learn in this book. These few key points are not very difficult—indeed, some students find them intuitively obvious.

Interpretations of Probability

probability
outcome

In statistics we usually define **probability** as "the expected relative frequency of a particular outcome." An **outcome** is the result of an experiment (or just about any situation in which the result is not known in advance, such as a coin coming up heads or it raining tomorrow). *Frequency* is how many times something happens. The *relative frequency* is the number of times something happens relative to the number of times it could have happened—that is, the proportion of times it happens. (A coin might come up heads 8 times out of 12 flips, for a relative frequency of 8/12, or 2/3.) **Expected relative frequency** indicates what you would expect to get in the long run, if you repeated the experiment many times. (In the case of a coin, in the long run you would expect to get 1/2 heads). This is called the **long-run relative-frequency interpretation of probability.**

expected relative frequency

long-run relative-frequency interpretation of probability

subjective interpretation of probability

We also use probability to communicate how certain we are that a particular thing will happen. This is called the **subjective interpretation of probability.** Suppose that you say there is a 95% chance that your favorite restaurant will be open tonight. You could be using a kind of relative frequency interpretation. This would imply that if you were to check whether this restaurant was open many times on days like today, on 95% of those days you would find it open. However, what you mean is probably more subjective: On a scale of 0% to 100%, you would rate your confidence that the restaurant is open at 95%. To put it another way, you would feel that a bet was fair that had odds based on a 95% chance of the restaurant's being open.

Which interpretation one adopts does not affect how probability is calculated. We introduced these concepts here for two reasons. First, we wanted to give you some deeper insight into the meaning of the term *probability,* which will be prominent throughout the rest of your learning about statistics, even if, as is so often the case, this deeper understanding does not take the form of set-in-stone dogma. Second, familiarity with both interpretations is crucial to understanding some of the deepest controversies in statistics—one of which we will introduce at the end of this chapter.

Calculating Probabilities

In statistical applications, probabilities are calculated as the proportion of successful outcomes—the number of possible successful outcomes divided by the number of all possible outcomes.

Consider the probability of getting heads when flipping a coin. There is one possible successful outcome (getting heads) out of two possible outcomes (getting heads or getting tails)—a probability of 1/2, or .5. In a throw of a single die, the probability of a 2 (or any other particular side of the die) is 1/6, or .17. This is because there can only be one particular successful outcome out of six possible outcomes. The probability of throwing a die and getting a number 3 or lower is 3/6, or .5. There are three possible successful outcomes (a 1, a 2, or a 3) out of six possible outcomes.

Box 5-2
Pascal Begins Probability Theory at the Gambling Table, Then Learns to Bet on God

Whereas in England statistics were used to keep track of death rates and to prove the existence of God (see Box 1-1), the French and the Italians developed statistics at the gaming table. In particular, there was the "problem of points"—the division of the stakes in a game after it has been interrupted. If a certain number of plays were planned, how much of the stakes should each player walk away with, given the percentage played so far?

The problem was discussed at least as early as 1494 by Luca Pacioli, a friend of Leonardo da Vinci. But it was unsolved until 1654, when it was presented to Blaise Pascal by the Chevalier de Méré. Pascal, a French child prodigy, attended meetings of the most famous adult French mathematicians and at 15 proved an important theorem in geometry. In correspondence with Pierre Fermat, another famous French mathematician, Pascal solved the problem of points and in so doing began the field of probability theory and the work that would lead to the normal curve.

By the way, not long after solving this problem, Pascal suddenly became as religiously devout as the English statisticians. He was in a runaway coach on a bridge and was saved from drowning only by the traces of the team breaking at the last possible moment. He took this as a warning to abandon his mathematical work in favor of religious writings and later formulated "Pascal's wager": that the value of a game is the value of the prize times the probability of winning it; therefore, even if the probability is low that God exists, we should gamble on the affirmative because the value of the prize is infinite, whereas the value of not believing is only finite worldly pleasure.

Reference: Tankard (1984).

Consider a slightly more complicated example. Suppose that a class has 200 people in it, and 30 are seniors. If you were to pick someone from the class at random, the probability of picking a senior would be 30/200, or .15. This is because there are 30 possible successful outcomes (getting a senior) out of 200 possible outcomes.

Range of Probabilities

Probabilities are proportions (the number of successful outcomes to the total number of possible outcomes). A proportion cannot be less than 0 or greater than 1. In terms of percentages, proportions range from 0% to 100%. Something that has no chance of happening has a probability of 0. Something that is certain to happen has a probability of 1. Notice that when the probability of an event is 0 the event is completely *impossible,* it cannot happen. But when the probability of an event is low, say 5% or even 1%, the event is *improbable* or *unlikely,* but not impossible.

Probabilities Expressed As Symbols

Probability is usually symbolized by the letter p. The actual probability number is usually given as a decimal, though sometimes fractions or percentages

p

are used. Thus a 50-50 chance is usually written as $p = .5$. But it could also be written as $p = 1/2$ or $p = 50\%$. It is also common to see probability written as being "less than" some number, using the "less than" sign (<). For example, "$p < .05$" means "the probability is less than 5%."

Probability Rules

As we noted above, our discussion only scratches the surface of probability. One of the topics we have not considered are the rules for computing probabilities involving multiple experiments or outcomes (for example, what is the chance of flipping a coin twice and both times getting heads?). These are called probability rules and they play an important role in the mathematical foundation of many aspects of statistics. However, knowing these probability rules is not necessary to understand the material covered in this book. Further, these rules are rarely used directly in analyzing results of psychology research. Nevertheless, you will occasionally see such procedures referred to in research articles. Thus, the two most widely mentioned probability rules are described in the Chapter Appendix.

Probability and the Normal Distribution

Until now we have mainly discussed probabilities of specific events that might or might not happen. We can also talk about a range of events that might or might not happen. The throw of a die coming out 3 or lower is an example. Other examples would be the probability of selecting someone on a city street who is between the ages of 30 and 40.

If you think of probability in terms of proportion of scores, probability fits in well with frequency distributions (see Chapter 1). Consider the frequency distribution shown in the histogram in Figure 5-8. Of the total of 50 scores, 10 are 7 or higher. If you were selecting people from this group at random, there would be 10 chances (possible successful outcomes) out of 50 (all possible outcomes) of selecting one that was 7 or higher. Thus, $p = 10/50 = .2$.

The normal distribution can also be thought of as a probability distribution. The normal curve is a frequency distribution in which the proportion of scores between any two Z scores is known. As we are seeing, the proportion of scores between any two Z scores is the same as the probability of selecting a score between those two Z scores. For example, the probability of a score

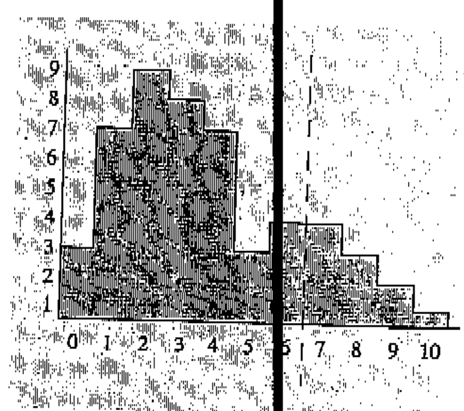


FIGURE 5-8
Frequency distribution (shown as a histogram) of 50 scores in which $p = .2$ (10/50) of randomly selecting a case with a score of 7 or higher.

being between the mean and a Z score of +1 (1 standard deviation above th[e] mean) is about 34%; that is, $p = .34$.

What we are saying may have been obvious all along. In a sense, it i[s] merely a technical point that the normal curve can be seen as either a fre[-]quency distribution or a probability distribution. We mention this only so tha[t] you will not be confused when we refer later to the probability of a score coming from a particular portion of the normal curve.

SAMPLE AND POPULATION

We are going to introduce you to some important ideas by thinking of beans. Suppose you are cooking a pot of beans and taste a spoonful to see if they are done. In this example, the pot of beans is a **population,** the entire set of things of interest. The spoonful is a **sample,** the part of the population about which you actually have information. This is illustrated in Figure 5-9.

population
sample

In psychology research, we typically study samples not of beans but of individuals. A sample might consist of 50 Canadian women who participate in a particular experiment; the population might be intended to be all Canadian women. In an opinion survey, 1,000 people might be selected from the voting-age population of a particular country and asked for whom they plan to vote. The opinions of these 1,000 people are the sample. The opinions of the larger voting public in that country, to which the pollsters hope to generalize their results, is the population (see Figure 5-10).[2]

Why Samples Are Studied Instead of Populations

As we have seen, researchers conduct studies to learn something about a population. Thus, their results would be most accurate if they could study the

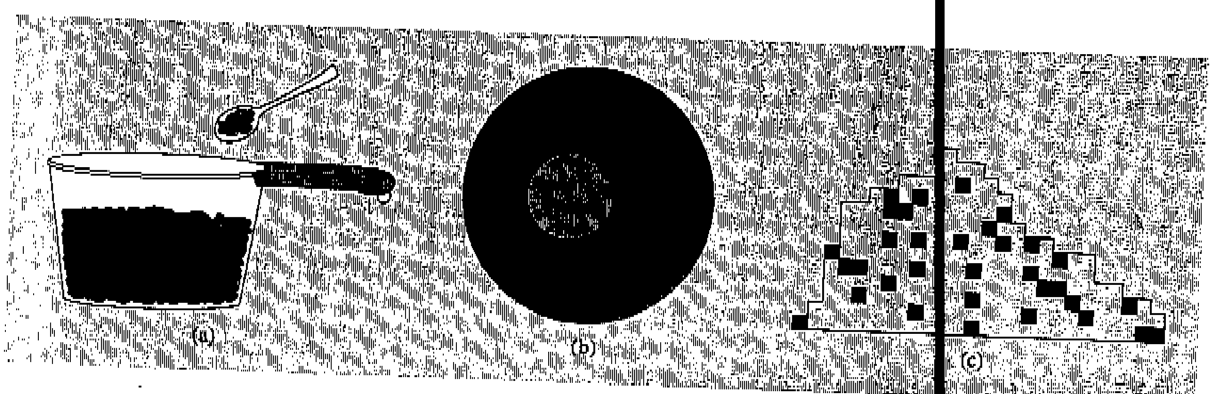
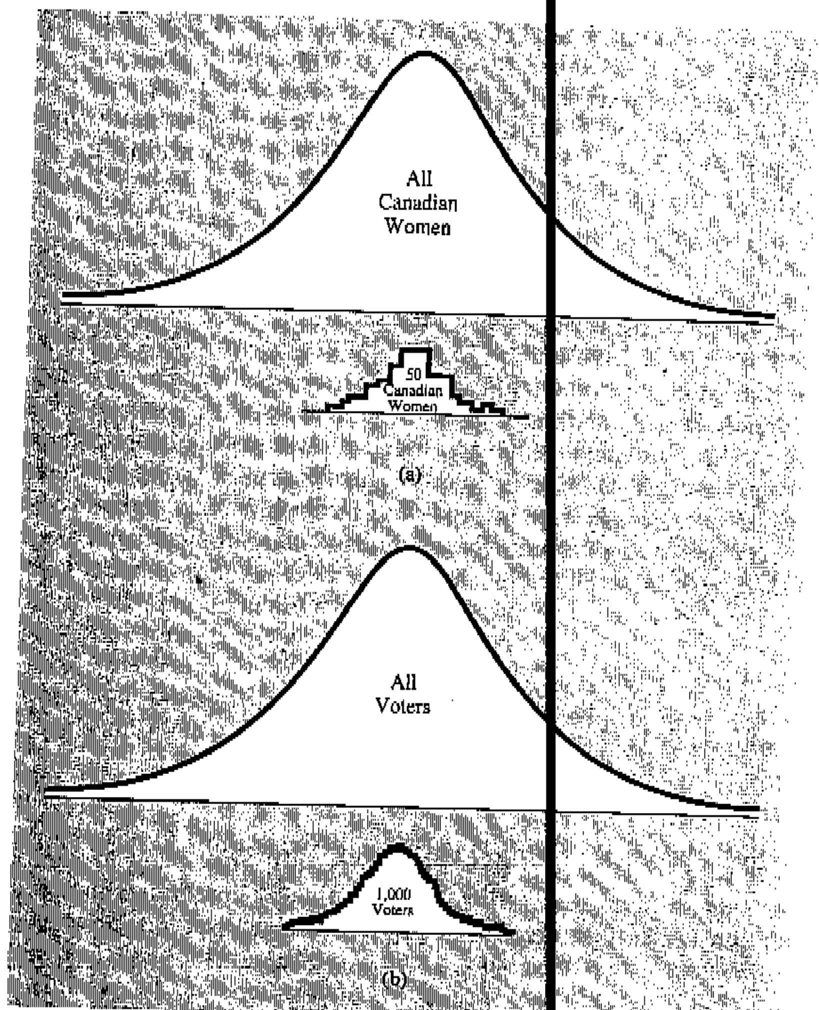


FIGURE 5-9
Populations and samples: In (a), the entire pot of beans is the population, the spoonful is a sample. [I]n (b), the entire larger circle is the population, the circle within it is the sample. In (c), the histogram is of the population and [t]he particular shaded scores together make up the sample.

[2]Strictly speaking, *population* and *sample* refer to a set of scores (numbers or measurements), not to the research participants measured. Thus, in the first example, the sample is really the scores of the 50 Canadian women, not the 50 women themselves; the population is really what the *scores* would be if all Canadian women were measured.

FIGURE 5-10
Additional examples of populations and samples. In (a), the population is the scores of all Canadian women, and a sample consists of the scores of the 50 particular Canadian women studied. In (b), the population is the voting preferences of the entire voting-age population in a particular country, and a sample consists of the voting preferences of the 1,000 voting-age people in that country who were surveyed.



entire population, rather than a subgroup from that population. However, in most research situations this is not practical. More important, usually the whole point of research is to be able to make generalizations or predictions about events beyond our reach. We would not call it scientific research if you tested your three cars to see which gets better gas mileage—unless you hoped to say something about the gas mileage of those models of cars in general. In other words, a researcher might conduct an experiment on the way in which people store words in short-term memory using 20 students as the participants in the experiment. But the purpose of the experiment is not to find out how these particular 20 students respond to the experimental condition. Rather, the purpose is to learn something about human memory under these conditions in general.

The strategy in almost all psychology research is to study a sample of individuals who are believed to be representative of the general population (or of some particular population of interest). More realistically, researchers try to study people who do not differ from the general population in any systematic way that would be expected to matter for that topic of research.

So in psychology research (and nearly all scientific research), the sampl is what is studied. The population is an unknown that researchers draw con clusions about on the basis of the sample. Most of what you learn in the res of this book is about the important work of drawing conclusions about popu- lations based on information from samples.

Methods of Sampling

There are so many ways of selecting a sample for a particular research project that we have provided a discussion of several of these methods in Appendix A (also see Box 5-3). Briefly, in most cases the ideal method of picking out a sample to study is called **random selection.** The researcher obtains a complete list of all the members of a population and randomly selects some number of them to study. An example of a random method of selection would be to put each name on a table tennis ball, put all the balls in a big hopper, shake it up, and have a blindfolded person select as many as are needed. (In practice, most researchers use a computer-generated list of random numbers. Just how computers or persons can create a list of truly random numbers is an interesting question in its own right that we examine in Box 15-1.)

random selection

It is important to distinguish truly random selection from what might be called **haphazard selection;** for example, just taking whoever is available or happens to be first on a list. When using haphazard selection, it is surprisingly easy accidentally to pick a group of people to study that is really quite different from the population as a whole. Consider a survey of attitudes about your statistics instructor. Suppose you collect your survey information from those who sit near you in class. Such a survey would be affected by all the factors influencing choice of seat, some of which have to do with precisely what you are studying—how much they like the instructor or the class. (Similarly, asking people who sit near you would result in opinions more like your own than a truly random sample would.)

haphazard selection

Unfortunately, it is only occasionally possible in psychology research to study a truly random sample. Much of the time, in fact, studies are conducted with whoever is willing or available to be a research participant. At best, as noted, a researcher tries to study a sample of individuals who are not systematically unrepresentative of the population in any known way. For example, suppose a study is about a process that is likely to differ for people of different age groups. In this situation, the researcher may attempt to include people of all age groups in the study. Alternatively, the researcher would be careful to draw conclusions only about the age group that was studied.

Statistical Terminology for Samples and Populations

The mean, variance, and standard deviation of a population are called **population parameters.** A population parameter is usually unknown and can only be estimated from what you know about a sample from that population. You don't taste all the beans, just the spoonful. "They're done" is an estimation about the whole pot.

opulation parameters

To keep this distinction in mind, population parameters are usually symbolized by Greek letters. The symbol for the mean of a population is μ, the Greek letter "mu." The symbol for the variance of a population is σ^2, and the symbol for its standard deviation is σ, the lowercase Greek letter "sigma."

TABLE 5-1
Population Parameters and Sample Statistics

	Population Parameter (Usually Unknown)	**Sample Statistic** (Computed From Known Data)
Basis:	Scores of entire population	Scores of sample only
Symbols:		
Mean	μ	M
Standard deviation	σ	SD
Variance	σ^2	SD^2

You won't see these symbols often, except while learning statistics. This is because, again, researchers seldom know the population parameters.

The mean, variance, and standard deviation you figure for the scores in a sample are called **sample statistics.** A sample statistic is computed from known information. Sample statistics are what we have been calculating all along and use the symbols we have used all along: M, SD^2, and SD. These various symbols are summarized in Table 5-1.

sample statistics

RELATION OF NORMAL CURVE, PROBABILITY, AND SAMPLE VERSUS POPULATION

In most research situations, as noted, we do not know the population parameters. However, we usually do assume that the shape of the population is roughly a normal curve. So researchers generally collect information from a sample in order to make probabilistic inferences about the parameters of a normally distributed population.

Consider an experiment to examine whether students learn more when studying all at once or when their studying is spread over a period of time. Sixty students are randomly selected to participate in the study. Half are randomly assigned to study all at once and half to study the same number of hours spread out over several weeks. At the end of the several weeks, both groups are given a test. The result is that there is a difference between the two groups in their mean scores on this test.

Now let's consider this experiment in terms of the language we have been using in this chapter. The group that studied all at once is a sample. This sample is intended to represent how students in general would perform if they were assigned to study all at once. That is, this sample represents a hypothetical population of students who are assigned to study all at once. The group that studied over time is another sample. This sample is intended to represent how students in general would perform if assigned to study over time. That is, this sample represents a hypothetical population of students who are assigned to study over time. The mean of each of the two groups studied is a sample statistic computed from the results of the experiment.

The populations represented by these samples do not even really exist. There is a general population of students, of course, but there is no population of students who have been assigned to these conditions (other than those in the experiment). We want to know about students who in the future might be given such instructions; that population is unknown. We usually assume that such unknown populations follow a normal curve. We make this assumption simply because most distributions in psychology follow a normal curve. However, we have no basis for making any assumptions about the mean and variance of those populations—these population parameters are unknown. Any conclusions we make about them must be based on the information in the sample statistics.

Finally, the question of interest is a question about probability. The thinking is a bit complex here—and we devote most of Chapter 6 to this. However, as a preview, consider the following logic: Suppose that the true means of the two populations (population parameters) were in fact the same. Under this supposition, how students study really doesn't affect amount learned. Nevertheless, when we did the experiment, the mean test scores of our two samples were different. Thus, given our supposition of no true difference between the populations, what is the probability that the means of our two samples could be as different as they actually are? If the probability is low, it seems unlikely our supposition of no difference between the populations was correct. Thus, we reject this supposition (of equal population means). If we reject this supposition of no difference between the populations, that leaves the conclusion that there is a difference between the populations. That is, this result supports the conclusion that the way students study really does affect how much they learn.

This logic may seem rather convoluted. It is. However, it is just such thinking about probabilities, samples, and populations that is the foundation of most inferential statistics in psychology. It is, in a nutshell, the logic of what is called "hypothesis testing," which we explore step-by-step in Chapter 6. You needn't ponder it now. We only introduced the ideas here to give you an idea of how the various elements covered in this chapter fit together in the kinds of statistical problems faced in real-life psychology research.

CONTROVERSIES AND LIMITATIONS

Basic as they are, the three topics we have introduced in this chapter—the normal curve, probability, and sample and population—are the subjects of considerable controversy. We shall explore a major controversy associated with each.

Is the Normal Curve Really So Normal?

We have said that real distributions in the world often approximate the normal curve very closely. The extent to which this is true is very important, and not merely because assuming a normal curve makes Z scores more useful. As you will see in later chapters, most of the statistical techniques that psychologists use assume that their samples come from populations that are normally distributed. The extent to which this is a reasonable assumption has been a longstanding source of debate. The predominant view has been that given the way psychological measures are developed, a bell-shaped distribution "is almost guaranteed" (Walberg, Strykowski, Rovai, & Hung, 1984, p. 107). Or as Hopkins and Glass (1978) put it, measurements in all disciplines are such good approximations to it that one might think "God loves the normal curve!"

However, there has been a persistent line of criticism about whether nature really packages itself so neatly. Micceri (1989) presented strong evidence that many measures commonly used in psychology do not yield scores tha are normally distributed "in nature." His study included achievement an ability tests (such as the SAT and the GRE) and personality tests (such as th MMPI). Micceri obtained data sets and examined the distributions of score of 440 psychological and educational measures that had been used on ver large samples. All of his data sets had samples of over 190 individuals, an the majority had samples of over 1,000 (14.3% even had samples of 5,000 t 10,293). Yet large samples were of no help. No distribution investigate passed all checks for normality (mostly Micceri looked for skewness, kurto sis, and "lumpiness"). Few measures had distributions that even came reason ably close to looking like the normal curve. Nor were these variation predictable: "The distributions studied here exhibited almost every conceiv able type of contamination" (p. 162), although some were more commo with certain types of tests. Micceri discusses many obvious reasons for thi nonnormality, such as "ceiling" or "floor" effects (see Chapter 2).

How much has it mattered that the distributions for these measures were so nonnormal? According to Micceri, it is just not known. And until more is known, the general opinion among psychologists will no doubt remain supportive of traditional statistical techniques, with the underlying mathematics based on the assumption of normal population distributions. What is the reason for this nonchalance in the face of findings such as Micceri's? It turns out that under most conditions in which they are used, the traditional techniques seem to give results that are reasonably accurate even when the formal requirement of a normal population distribution is not met (e.g., Sawilowsky & Blair, 1992). In this book, we generally adopt this majority position favoring the use of the traditional techniques in all but the most extreme cases. But you should be aware that a vocal minority of psychologists disagrees. Some of the alternative statistical techniques they favor (ones that do not rely on assuming a normal distribution in the population) are presented in Chapter 15.

Galton, one of the major pioneers of statistical methods (recall Box 3-1), said of the normal curve, "I know of scarcely anything so apt to impress the imagination. . . . [It] would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wild confusion" (1889, p. 66). Ironically, it may be true that in psychology, at least, it truly reigns in pure and austere isolation, with no even close-to-perfect real-life imitators.

What Does Probability Really Mean?

We have already introduced the major controversy in the area of probability theory as it is applied to statistics in psychology, the dispute between the long-term relative-frequency interpretation and the subjective degree-of-belief interpretation. In most cases, though, it really does not much matter which interpretation is used; the statistics are the same. But among the minority of theorists who favor the subjective interpretation, some hold a rather critical view of the mainstream of statistical thinking. In particular, they have advocated what has come to be called the "Bayesian approach" (for example, see Phillips, 1973). The approach is named after Thomas Bayes, an early-18th-century nonconformist English clergyman who developed a probability theorem appropriately known as "Bayes' theorem."

Bayes' theorem itself can be proved mathematically and is not controversial. However, its applications in statistics are hotly debated. The details of the approach are beyond the scope of an introductory text, but the main point of dispute can be made clear here: According to the Bayesians, science is about conducting research in order to adjust our preexisting beliefs in light of evidence we collect. Thus, conclusions drawn from an experiment are always in the context of what we believed about the world before doing the experiment. The mainstream view, by contrast, says that it is better not to make any assumptions about prior beliefs. We should just look at the evidence as it is, judging whether the experiment has shown any reliable effects at all (or no effect whatsoever). Some statisticians in the mainstream do acknowledge that the Bayesian description of science may be more accurate. However, they are uncomfortable with using Bayesian methods in statistical computations in research practice, because adopting them would mean that the conclusion drawn from each study would depend too heavily on the subjective belief of the particular scientist conducting the study. Thus, the same experimental results from different scientists could lead to different conclusions.

The Bayesian approach represented a lively (though never majority) movement in psychological statistics during the 1960s and 1970s. Since then it has become much less prominent as a movement, at least under this banner. Nevertheless, many of the issues raised by this dispute remain important in new guises (Games, 1988; Gigerenzer & Murray, 1987; Leventhal & Huyn, 1996; Prentice & Miller, 1992).

Sample and Population

Most of the statistical procedures you learn in the rest of this book are based on the assumption that the sample studied is a random sample of the population. As we pointed out, however, this is rarely the case in psychology research. Most often, our samples include whatever individuals are available to participate in our experiment—meaning that most studies are done with college students, volunteers, convenient laboratory animals, and the like.

Some psychologists are concerned about this problem and have suggested that researchers need to use different statistical approaches that make generalizations only to the kinds of people that are actually being used in the study.[3] For example, these psychologists would argue that if our sample has a particular nonnormal distribution, we should assume that we can generalize only to a population with the same particular nonnormal distribution. We will have more to say about their suggested solutions in Chapter 15.

[3]Frick (in press) argues that in most cases psychology researchers should not think in terms of samples and populations at all. Rather, he argues, researchers should think of themselves as studying processes. An experiment examines some process in a group of individuals. Then the researcher evaluates the probability that the pattern of results could have been caused by chance factors. For example, the researcher examines whether a difference in means between an experimental and a control group could have been caused by factors other than by the experimental manipulation. Frick claims that this way of thinking is much closer to the way researchers actually work, and argues that it has various advantages in terms of the subtle logic of inferential statistical procedures. It will be interesting to see the reaction to Frick's proposal. In any case, following the more standard approach (as taught in this book) yields exactly the same results and is consistent with the way most psychologists understand statistical reasoning.

Sociologists, as compared to psychologists, are much more concerne about the representativeness of the groups they study. Studies reported i sociology journals (or in sociologically oriented social psychology journals are much more likely to use formal methods of random selection and larg samples or at least to address the issue in their articles.

Why are psychologists more comfortable with using not clearly randon samples? The primary reason is that they are mainly interested in the *rela tionships* among variables. If in one population an increase in X is associate with an increase in Y, this relationship should probably hold in other popula tions. This relationship should hold even if the actual levels of X and Y diffe from population to population. Suppose that a researcher conducts the exper iment we used as an example in Chapters 3 and 4, testing the relation of num ber of exposures to a list of words to number of words remembered. Suppose further that this study is done with college students and that the result is tha the greater the number of exposures, the greater the number of words remembered. The actual number of words remembered from the list might well be different for people other than college students. For example, chess masters (who probably have highly developed memories) may recall more words; people who have just been upset may recall fewer words. However, even in these groups, we would expect that the more times exposed to the list, the more words will be remembered. That is, the *relation* of number of exposures to number of words recalled will probably be about the same in each population.

In sociology, the representativeness of samples is much more important. This is because sociologists are more concerned with the actual mean and variance of a variable in a particular society. Thus, a sociologist might be interested in the average attitude towards older people in the population of a particular country. For this purpose, how sampling is done is extremely important.

NORMAL CURVES, PROBABILITIES, SAMPLES, AND POPULATIONS IN RESEARCH ARTICLES

The material covered in this chapter is primarily used as a foundation for understanding the topics covered in subsequent chapters. It is rarely discussed explicitly in research articles (except articles about methods or statistics). Occasionally, you will see the normal curve mentioned, in the context of describing the scores on a particular variable. (We say more about this, and give some examples from published articles, in Chapter 15. In that chapter we consider circumstances in which the scores do not follow a normal curve.)

Probability is also rarely discussed directly, except in the context of statistical significance, a topic we mentioned briefly in Chapter 3. In almost any article you look at, the Results section will be strewn with descriptions of various methods associated with statistical significance, followed by some expression such as "$p < .05$" or "$p < .01$." The p refers to probability, but the probability of what? That is the main topic of our discussion of statistical significance in Chapter 6.

Finally, you will occasionally see a brief mention of the method of selecting the sample from the population. For example, Altman, Levine, Howard, and Hamilton (1997) conducted a telephone survey of the attitudes of the

U.S. adult public towards tobacco farmers. In the Method section of their article, they explained that their respondents were "randomly selected from a nationwide list of telephone numbers" (p. 117). Thus, Altman et al. specified both the listing they used for the population (the nationwide list of phone numbers) and the method they used (random selection) to obtain their sample. It is worth noting, however, that in such surveys the response rate of those telephoned is usually far from 100%. In this example, they obtained interviews with 47% of those they called. So even though they used random selection to contact potential members of their sample, the sample itself was not random. The sample overrepresents whatever characteristics make a person available and willing to respond to a telephone survey.

SUMMARY

The scores on many variables in psychology research approximately follow a bell-shaped, symmetrical, unimodal shape called the normal curve. Because the shape of this curve follows an exact mathematical formula, there is a specific percentage of scores between any two points on a normal curve.

Important figures to remember about a normal curve are that 34% of the scores are between the mean and 1 standard deviation above the mean, and 14% between 1 and 2 standard deviations above the mean.

A normal curve table gives the percentage of scores between the mean and any particular positive Z score. Using this table, and knowing that the curve is symmetrical and that 50% of the scores are above the mean, you can determine the percentage of scores above or below any particular Z score. You can also use the table to determine the Z score for the point at which a particular percentage of scores begins.

Most psychology researchers consider the probability of an event to be its expected relative frequency. However, some think of probability as the subjective degree of belief that the event will happen. Probability is usually calculated as the proportion of successful outcomes to total possible outcomes. It is symbolized by p and has a range from 0 (event is impossible) to 1 (event is certain). The normal curve gives the probabilities of scores being within particular ranges of values.

A sample is an individual or group that is studied, usually as representative of a larger group, or population, that cannot be examined in its entirety. Ideally, the sample is selected from a population using a strictly random procedure. The mean, variance, and so forth of a sample are called sample statistics. When of a population, they are called population parameters and are symbolized by Greek letters—μ for mean, σ^2 for variance, and σ for standard deviation.

Most of the techniques you learn in the rest of this book make probabilistic inferences to draw conclusions about populations based on information from samples. In this process, populations are usually assumed to be normally distributed.

There are controversies relating to each of the major topics. One question is about whether normal distributions are truly typical of the populations of scores for the variables we study in psychology. Another debate, raised by advocates of a "Bayesian" approach to statistics, is whether we should explicitly construct our statistical procedures to take the researcher's initial subjective expectations into account. Finally, the representativeness of the samples that psychologists use, which are typically not obtained through strict random selection, has been contested—though there are also reasons to think that for the topics most psychologists study, this may not matter very much.

Research articles rarely discuss normal curves (except briefly when the distribution at hand seems not to be normal) or probability (except in the context of significance testing, described beginning in Chapter 6). However, procedures of sampling, particularly when the study is a survey, are usually described, and the representativeness of a sample when random sampling could not be used may be discussed.

Key Terms

expected relative frequency
haphazard selection
long-run relative-frequency interpretation of probability
normal curve
normal curve table
normal distribution
outcome
population
population parameters
probability (p)
random selection
sample
sample statistics
subjective interpretation of probability
μ
σ
σ^2

Practice Problems

These problems involve computation (with the assistance of a calculator). Most real-life statistics problems are done on a computer. But even if you have a computer, do these by hand to ingrain the method in your mind.

For practice in using a computer to solve statistical problems, refer to the computer section of each chapter of the *Student's Study Guide and Computer Workbook* that accompanies this text.

All data are fictional (unless an actual citation is given).

Answers to Set I problems are given at the back of the book.

SET I

1. Suppose that the people living in a particular city were found to have a mean score of 40 and a standard deviation of 5 on a measure of concern about the environment. Assume that these concern scores are normally distributed. Approximately what percentage of people have a score (a) above 40, (b) above 45, (c) above 30, (d) above 35, (e) below 40, (f) below 45, (g) below 30, and (h) below 35? What is the minimum score a person has to have to be in the top (i) 2%, (j) 16%, (k) 50%, (l) 84%, and (m) 98%? (Use the 50%–34%–14% figures for this problem.)

2. A psychologist has been studying eye fatigue using a particular measure which she administers to students after they have worked for 1 hour writing on a computer. On this measure, she has found that the distribution follows a normal curve. What percentage of students has Z scores (a) below 1.5, (b) above 1.5, (c) below −1.5, (d) above −1.5, (e) above 2.10, (f) below 2.10, (g) above .45, (h) below −1.78, and (i) above 1.68?

3. Assuming a normal curve, (a) if a person is in the top 10% of their country on mathematics ability, what is that person's Z score? (b) If the person is in the top 1%, what would be the Z score?

4. Consider a test of coordination that has a normal distribution, a mean of 50, and a standard deviation of 10. How high a score would a person need to be in the top 5%? Explain your answer to someone who has never had a course in statistics.

5. The following numbers of individuals in a company received special assistance from the personnel department last year:

Drug/alcohol	10
Family crisis counseling	20
Other	20
Total	50

If you were to select someone at random from the records for last year, what is the probability that the person would be in the (a) drug/alcohol, (b) family, (c) drug/alcohol or family, (d) any category except "Other," (e) any of the three categories?