APPENDIX

Statistics for Psychology

The Law of Large Numbers

If you sat down and started picking cards from a deck, you would expect to pick as many red cards as black cards over the long run. But *only* over the long run. You would not be surprised if you picked just two cards and they both turned out to be red, but you *would* be surprised if you picked just 20 cards and they all turned out to be red. Your intuition tells you that when the number of cards you pick is small, you really can't expect your hand to have the same proportion of red and black cards as does the full deck.

Your intuition is exactly right. Remember from Chapter 2 that a *population* is the complete collection of objects or events that might be measured, and a *sample* is the partial collection of objects or events that is measured. In this case, the full deck is a population, and the cards in your hand are a sample of that population. Your intuition about the cards is captured by the **law of large numbers,** which states that as sample size increases, the attributes of the sample more closely reflect the attributes of the population from which the sample was drawn. In plain English, the more cards you pick, the more likely it is that half the cards in your hand will be red and half will be black.

Precisely the same logic informs the methods of psychology. For example, if we wanted to know how happy people are in Florida, we would begin with an operational definition of happiness. For the sake of simplicity, we might define happiness as a person's belief about his or her own emotional state. Then we'd develop a way to measure that belief, for example, by asking the person to make a checkmark on a 10-point rating scale. If we used this measure to measure the happiness of just one Floridian, our lone observation would tell us little about the happiness of the roughly 19 million people who actually live in that state. However, if we were to measure the happiness of a hundred Floridians, or a thousand Floridians, or even a million Floridians, the average of our measurements would begin to approximate the average happiness of all Floridians. The law of large numbers suggests that as the size of our sample increases, the average happiness of the people in our sample becomes a better approximation of the average happiness of the people in the population.

law of large numbers States that as sample size increases, the attributes of the sample more closely reflect the attributes of the population from which the sample was drawn.

Averaging

If we have chosen a representative and sufficiently large sample, the average happiness of that sample can tell us about the average happiness of the population from which it was drawn. But the average cannot tell us about the happiness of particular individuals in that population. For example,

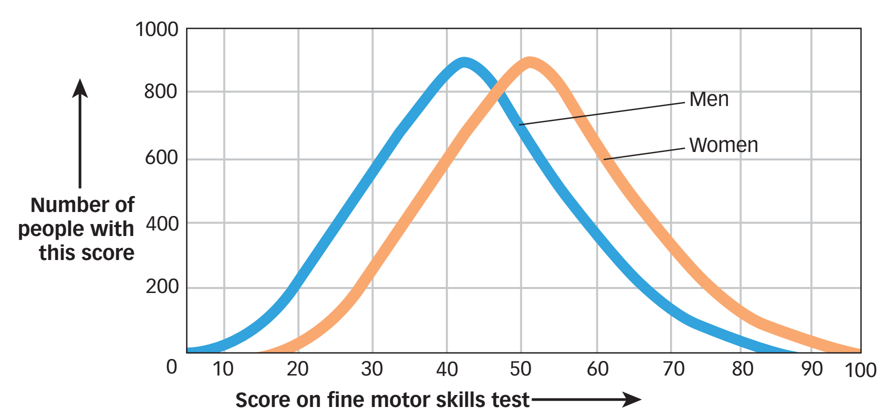
AP PHOTO/OTTO BALLON MIERNY/CTK

On average, men have more upper-body strength than women, but there are still many women with more upper-body strength than many men.

when psychologists claim that women have better fine motor skills than men (and they do), or that men have better spatial ability than women (and they do), or that children are more suggestible than adults (and they are), their claims are not true—and are not meant to be true—of every individual in these populations. Rather, when psychologists say that women have better fine motor skills than men, they mean that when the fine motor skills of a large sample of women and men are measured, the average of the women's measurements is reliably higher than the average of the men's.

FIGURE A.1 illustrates this point with hypothetical observations that are arranged in a pair of **frequency distributions**, which are graphic representations of the measurements of a sample that are arranged by the number of times each measurement as observed. These frequency distributions display every possible score on a fine motor skills test on the horizontal axis and display the number of times (or the frequency with which) each score was observed among a sample of men and women on the vertical axis. A frequency distribution can have any shape, but it commonly takes the shape known as a normal distribution (sometimes also called a bell curve). A **normal distribution** is a frequency distribution in which most measurements are concentrated around the mean and fall off toward the tails, and the two sides of the distribution are symmetrical. As you can see in FIGURE A.1, normal distributions are symmetrical (i.e., the left half is a mirror image of the right half), have a peak in the middle, and trail off at either end. Most scores can be found toward the center of a normal distribution, with fewer scores at the extremes. In fact, the point at the very center of a normal distribution is where you'll find the average.

FIGURE A.1 **Frequency Distributions** *This graph shows the hypothetical scores of a sample of men and women who took a test of fine motor skills. The scores are represented along the horizontal axis, and the frequency of each score is represented along the vertical axis. As you can see, the average score of women is a bit higher than the average score of men. Both distributions are examples of normal distributions.*



Descriptive Statistics

A frequency distribution depicts every measurement in a sample and thus provides a full and complete picture of that sample. But like most full and complete pictures, it is a terribly cumbersome way to communicate information. When we ask a friend how she's been, we don't want her to show us a graph depicting her happiness on each day of the previous six months. Rather, we want a brief summary statement that captures the essential information that such a graph would provide—for example, "I'm doing pretty well" or "I've been having some ups and downs lately." In psychology, brief summary statements that capture the essential information from a frequency distribution are called *descriptive statistics*. There are two important kinds of descriptive statistics:

- *Descriptions of central tendency* are summary statements about the value of the measurements that lie near the center or midpoint of a frequency distribution. When a friend says that she has been "doing pretty well," she is describing the central tendency (or approximate location of the midpoint) of the frequency distribution of her happiness measurements. The three most common descriptions of central tendency are the **mode** (the value of the most frequently observed measurement), the **mean** (the average value of all the measurements), and the **median** (the value that is greater than or equal to the values of half the measurements and less than or equal to half the values of the measurements). In a normal distribution, the mean, median, and mode are all the same value, but when the distribution departs from normality, these three descriptive statistics can differ. FIGURE A.2 shows how each of these descriptive statistics is calculated.
- *Descriptions of variability* are statements about the extent to which the measurements in a frequency distribution differ from each other. When a friend says that she has been having some "ups and downs" lately, she is offering a brief summary statement that describes how the measurements in the frequency distribution of her happiness scores over the past six months tend to differ from one another. A mathematically simple description of variability is the **range**, which is the value of the largest measurement minus the value of the smallest measurement (FIGURE A.2). There are several other common descriptions of variability, such as the *variance* and the *standard deviation*, but all such descriptions give us a sense of how similar or different the scores in a distribution tend to be.

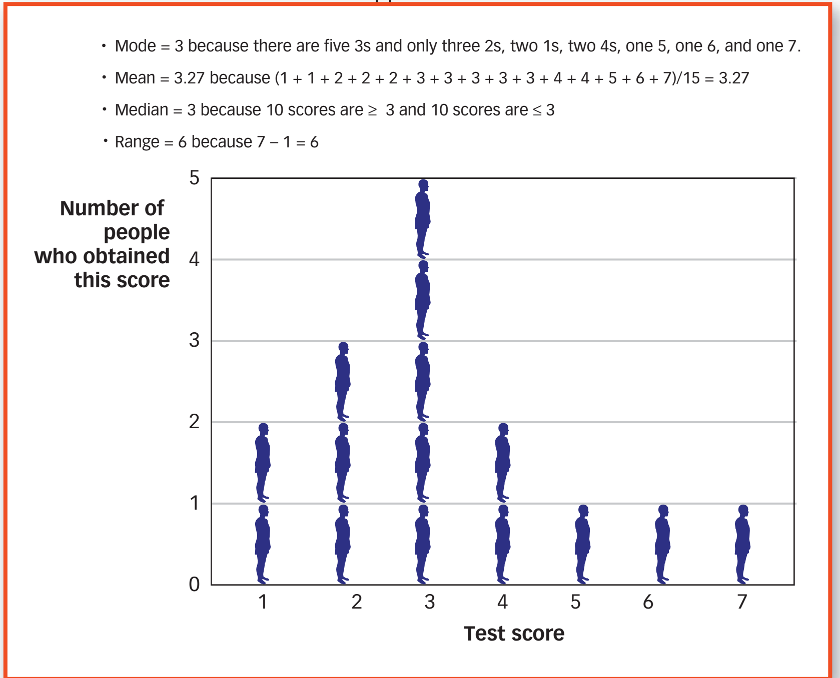


FIGURE A.2 **Some Descriptive Statistics** *This frequency distribution shows the scores of 15 individuals on a seven-point test. Descriptive statistics include measures of central tendency (such as the mean, median, and mode) and measures of variability (such as the range).*

Measuring Correlation

Every correlation can be described in two equally reasonable ways. A positive correlation describes a relationship between two variables in "more-more" or "less-less" terms. When we say that *more age* is associated with *more height* or that *less age* is associated with *less height*, we are describing a positive correlation. A negative correlation describes a relationship between two variables in "more-less" or "less-more" terms. When we say that *more cholesterol* is associated with *less longevity* or that *less cholesterol* is associated with *more longevity*, we are describing a negative correlation. How we choose to describe any particular correlation is usually just a matter of simplicity and convenience.

Of course, just because two variables are correlated doesn't mean that every single individual follows the "more-more" or "more-less" rule. There is a correlation between age and height, and most children are shorter than most adults. But not in every case. There are *some* tall kids and some *short* adults. The **correlation coefficient** is a measure of the direction and strength of a correlation, and it is symbolized by the letter *r* (as in "relationship"). Like most measures, the correlation coefficient has a limited range. If you were to measure the number of hours of sunshine per day in your hometown, that measure would have a range of 24 because it could only have a value from 0 to

frequency distributions Graphic representations of the measurements of a sample that are arranged by the number of times each measurement was observed.

normal distribution A frequency distribution in which most measurements are concentrated around the mean and fall off toward the tails, and the two sides of the distribution are symmetrical.

mode The value of the most frequently observed measurement.

mean The average value of all the measurements.

median The value that is greater than or equal to the values of half the measurements and less than or equal to half the values of the measurements.

range The value of the largest measurement minus the value of the smallest measurement.

correlation coefficient A measure of the direction and strength of a correlation, and it is symbolized by the letter *r* (as in "relationship").

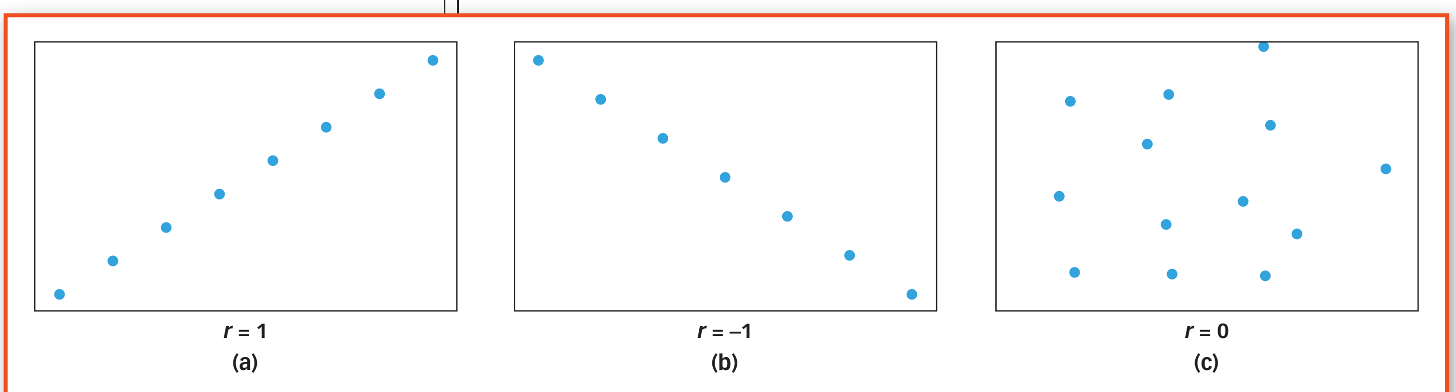
When children line up by age, they also tend to line up by height. The pattern of variation in age (from youngest to oldest) is synchronized with the pattern of variation in height (from shortest to tallest).

PETER TURNLEY/CORBIS

24. Values such as −7 and 36.8 would be meaningless. Similarly, the value of *r* can range from −1 to 1, and numbers outside that range are meaningless. What, then, do the numbers *inside* that range mean?

- When $r = 1$, the relationship between the variables is called a *perfect positive correlation*, which means that every time the value of one variable increases by a certain amount, the value of the second variable always increases by a certain amount, too. If every increase in age of one year was associated with an increase in height of, say, three inches, then age and height would be perfectly positively correlated (**FIGURE A.3a**).
- When $r = -1$, the relationship between the variables is called a *perfect negative correlation*, which means that as the value of one variable increases by a certain amount, the value of the second variable *decreases* by a certain amount, and this happens without exception. If every increase in age of one year were associated with a decrease in height of, say, one inch, then age and height would be perfectly negatively correlated (**FIGURE A.3b**).
- When $r = 0$, there is no systematic relationship between the variables, which are said to be *uncorrelated*. This means that the pattern of variation of one variable is not synchronized in any way with the pattern of variation of the other. If increases in age of one year were sometimes associated with a large increase in height, sometimes with a small increase in height, and sometimes with no increase at all—or even a decrease—then age and height would be uncorrelated (**FIGURE A.3c**).

FIGURE A.3 **Three Kinds of Correlations** *This figure illustrates pairs of variables that have (a) a perfect positive correlation ($r = 1$), (b) a perfect negative correlation ($r = -1$), and (c) no correlation ($r = 0$).*



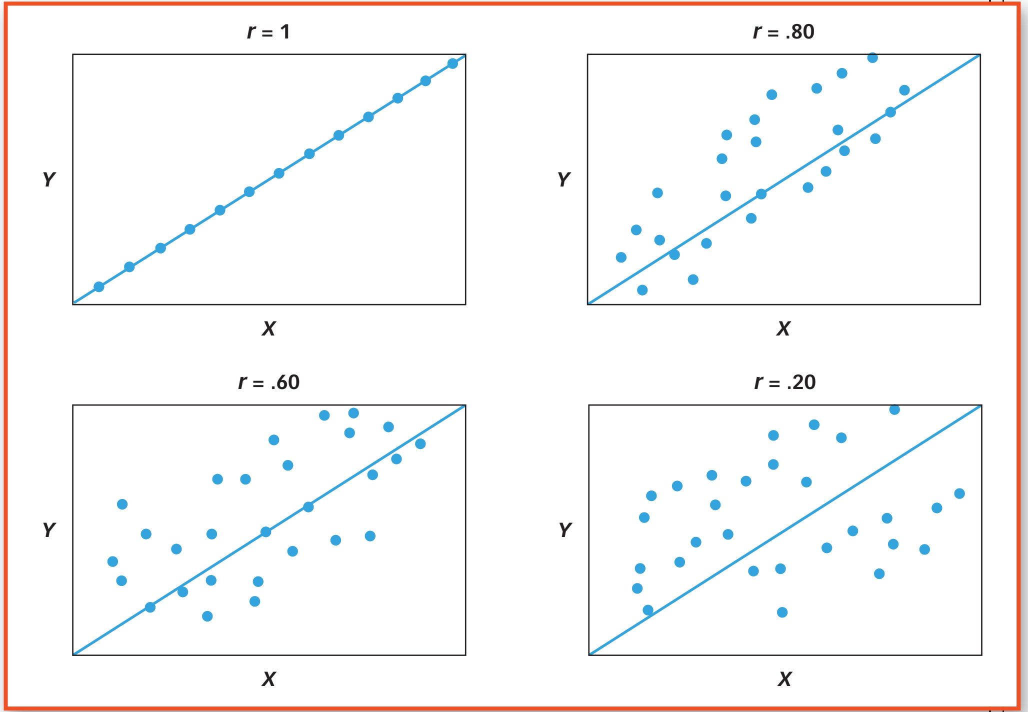


FIGURE A.4 **Positive Correlations of Different Strengths** *These graphs represent different degrees of positive correlation between two variables. Scores that are on the line adhere strictly to the rule X = Y. The more exceptions there are to this rule, the weaker the correlation is.*

Perfect negative and positive correlations, such as those in **FIGURES A.3a** and **A.3b**, are extremely rare in real life. Age and height are correlated, but they are imperfectly correlated; that is, exceptions to the "more-more" rule clearly do exist. The more exceptions there are, the closer to zero *r* will be. **FIGURE A.4** shows four cases in which two variables are positively correlated but have different numbers of exceptions, and as you can see, the number of exceptions changes the value of *r* quite dramatically. Two variables can have a perfect correlation ($r = 1$), a strong correlation (e.g., $r = .80$), a moderate correlation (e.g., $r = .60$), or a weak correlation (e.g., $r = .20$). The correlation coefficient, then, is a measure of both the *direction* and *strength* of the relationship between two variables. The sign of *r* (positive or negative) tells us the direction of the relationship, and the absolute value of *r* (between 0 and 1) tells us about the number of exceptions to the rule.

Statistical Significance

The correlation coefficient *r* gives us a way to assess the strength of a correlation between two variables. But it tells us nothing about the causal relationship between these two variables. How do we tell if a change in one variable is actually *causing* the change in another? In Chapter 2, you learned that the way to do this was by conducting an experiment that involves manipulating an independent variable (thereby creating an experimental group and a control group) and then measuring a dependent variable. If the average measurement in the two groups differs, then you can conclude that this difference was caused by the manipulation.

There's just one small problem with this conclusion, and that's that every once in a while, differences are caused by chance. How can we tell if the differences we observed in an experiment were caused by chance? Unfortunately, we can't tell for sure. But we can calculate the *odds* that they were caused by chance. A sophisticated mathematical technique allows psychologists to calculate *p*, which is the likelihood that a measured difference between two groups was caused by chance. If $p < .05$, then the odds are less than 5% that differences between the experimental and control groups were caused by chance. Psychologists do not accept the results of an experiment unless $p < .05$, and when it is, the result is said to be *statistically significant*.