Key Question Chapter Outline

How Do Psychologists Develop New Knowledge?

The Five Steps of the Scientific Method
Types of Psychological Research
Sources of Bias in Research (or Anywhere Else)
Ethics in Research
Questions Science Cannot Answer

How Do We Make Sense of the Data?

Developing Your Own Survey
Organizing the Data
Describing the Data with Descriptive Statistics
Correlation: A Relationship Between Two Variables
Making Inferences with Inferential Statistics

CORE CONCEPTS

Psychologists, like researchers in all other sciences, use the scientific method to test their ideas empirically.

Researchers use statistics for two major purposes: (1) descriptively to characterize measurements made on groups or individuals and (2) inferentially to judge whether those measurements are the result of chance.

Psychology in Your Life

Getting in Deeper

Whatever your major, consider getting a student membership in a professional organization and reading some of the journals.

Statistics in Politics

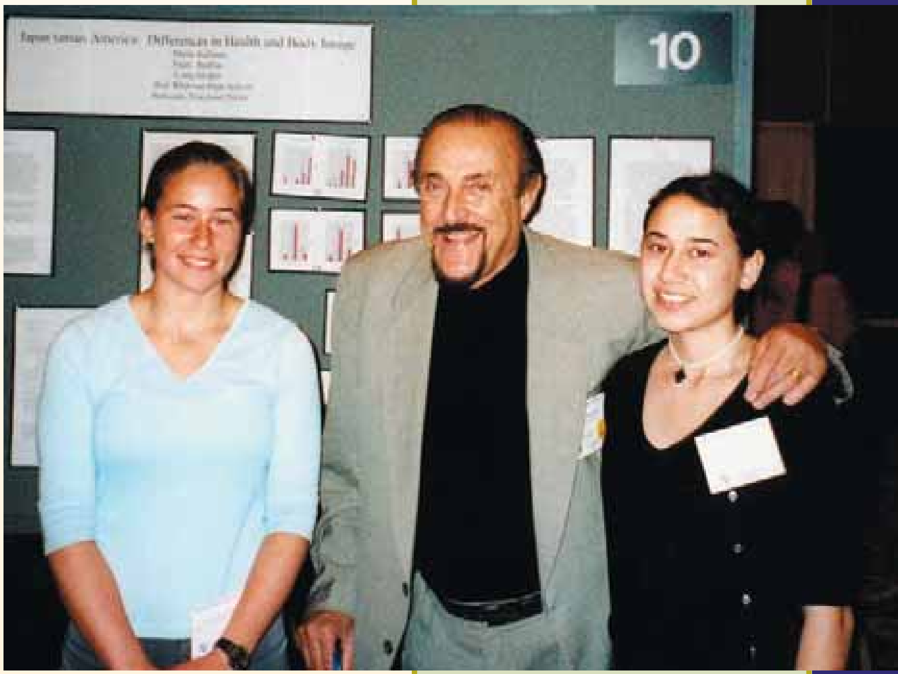
Politicians rely on statistics to measure the responses to their platform, ideas, and job performance.

USING PSYCHOLOGY TO LEARN PSYCHOLOGY
Research in Practice

Chapter

2

Research Methods



YUMI AND MARIA WERE TWO teenage girls from similar suburban backgrounds who grew up together and attended the same schools. Intrigued by their psychology class discussions, Yumi, of Japanese descent, and Maria, of European/American descent, began contemplating whether the differences in their heritage affected their views on personal issues.

The teens were determined to find out if Japanese and American teenagers viewed body image and health issues differently. Specifically, they wanted to know if either group perceived themselves as heavier or thinner; whether or not there were differences in the perceptions of body image between the two groups; and whether there were any differences between body image and health among the groups.

Admittedly a complicated quest, the teens first looked around for existing information to answer these questions. Finding none, the girls decided to conduct their own research. So, how did two teenagers from Washington, D.C. go about finding answers to their questions? They used a scientific approach. Beginning with their initial discussions, the identification of their questions, the review of available research, and on through the creation of a survey, analysis of the results, and a review of their methodology, Yumi and Maria employed the basic methods of science—the same principles and processes psychologists use everyday to answer their own questions, test their own theories, and gain knowledge. The scientific method, whether conducted by teenagers to answer personal questions about their peers, or conducted by psychologists, is our focus in this chapter.

KEY QUESTION

HOW DO PSYCHOLOGISTS DEVELOP NEW KNOWLEDGE?

As early as 1880, psychologists were challenging the claims of spiritualists and psychics (Coon, 1992). But even today, psychology continues to dispute the unfounded claims of *pseudoscience,* which seem to blossom faster than they can be nipped in the bud. Modern sources of questionable psychology include practitioners of astrology, palmistry, graphology, biorhythm analysis, and any number of psychics, seers, and prophets who claim to have special insights into people's personalities and to be able to predict their futures.

But what makes psychology different from the pseudoscientific approaches to understanding people? Answer: None of the pseudosciences has survived trial by the *scientific method,* which is a way of rigorously testing ideas against objective observations. Instead, pseudoscience is based on mere speculation and anecdote—and on human gullibility.

You might think this a snobbish view for psychologists to take. Why can't we make room for many different approaches to the understanding of people? In fact, we do. Psychologists have no problem with sociology, anthropology, and psychiatry, for example, as partners in the enterprise of understanding people. Psychologists reject only those approaches that claim to have "evidence" but offer only anecdotes and testimonials.

What, then, makes psychology a real science? Again, it's the *method.* As our Core Concept for this section says:

CORE CONCEPT

Psychologists, like researchers in all other sciences, use the scientific method to test their ideas empirically.

What is this marvelous method? Simply put, the **scientific method** is a process for putting ideas to an objective pass–fail test. At the heart of this testing procedure is **empirical investigation,** the collecting of objective information firsthand by making careful measurements based on direct experience. Literally, *empirical* means "experience based"—as contrasted with speculation based solely on faith, hope, authority, or common sense. To investigate a question empirically is to collect evidence yourself, rather than relying solely on a logical argument or appealing to the opinion of "experts." Ultimately, a main goal of psychological science is to develop explanations for behavior and mental processes—explanations based on solid empirical studies. We call these explanations *theories.*

CONNECTION: CHAPTER 6

Social learning is acquiring a new behavior by watching others and seeing how they are rewarded and punished for their behavior.

In brief, a **theory** is a testable explanation for a set of facts or observations (Kerlinger, 1985; Kukla, 1989). Please note that this definition may be quite different from the way you customarily use the term. In everyday language, "theory" can mean "wild speculation" or a mere "hunch"—an idea that has no evidence to support it. "It's only a theory," people may say. But *theory* means something quite different to a scientist. The essence of a scientific theory is its power to explain the facts and its ability to be tested objectively. Some theories have a great deal of evidence to support them, while others are highly speculative. Examples of well-supported theories include Einstein's theory of relativity, the germ theory of disease, Darwin's theory of natural selection, and, in psychology, social learning theory (which we will discuss in Chapter 6).

To illustrate the scientific method in action, we would remind you how Dr. Pfungst put Clever Hans to the test. But to take a more recent example, let's look at a simple and elegant psychological experiment published in the *Journal of the American Medical Association* by . . . a fourth grader (Rosa et al., 1998)! Meet Emily Rosa of Loveland, Colorado. Emily's school science project, it

■ **Scientific method** A five-step process for empirical investigation of a hypothesis under conditions designed to control biases and subjective judgments.
■ **Empirical investigation** An approach to research that relies on sensory experience and observation as research data.
■ **Theory** A testable explanation for a set of facts or observations. In science, a theory is *not* just speculation or a guess.

turned out, challenged a widely held belief in the power of *therapeutic touch* (TT).

In the early 1990s, TT was touted as a medical therapy, and Emily's mother, a nurse, had explained to her how TT practitioners attempted to promote healing by moving their hands over the patient's body without directly touching it. In doing so, they believed that they were detecting and manipulating an energy field radiating from the body. These practitioners claimed they could use TT to treat a wide range of medical and psychological problems—from colic to cancer and arthritis to depression (Gorman, 1999). So effective was it believed to be that the technique was being taught in more than 100 colleges and universities in 75 countries and used by nurses in at least 80 U.S. hospitals.

But did it really work, or was it just another example of flawed common sense? Emily Rosa suspected that TT practitioners were really detecting their own beliefs and expectations, rather than a "human energy field." So she put their claims to a simple experimental test, the details of which we will use to illustrate the scientific method.

Emily Rosa's experiment

The Five Steps of the Scientific Method

Testing any scientific assertion requires five steps. (See Figure 2.1.) These steps are essentially the same whether the study involves psychology, biology, chemistry, astronomy, or any other scientific discipline. Thus it is the *method* that makes these fields scientific, not their subject matter. Ideally, a researcher (such as Emily Rosa) who follows the scientific method will proceed as follows.

Developing a Hypothesis The first step calls for coming up with a testable idea, or prediction. Scientists call this prediction a **hypothesis.** The term literally means "little theory" because it often represents only one piece of a larger

Hypothesis A statement predicting the outcome of a scientific study; a statement describing the relationship among variables in a study.

1. Developing a hypothesis

2. Performing a controlled test

3. Gathering objective data

4. Analyzing the results

5. Publishing, criticizing, and replicating the results

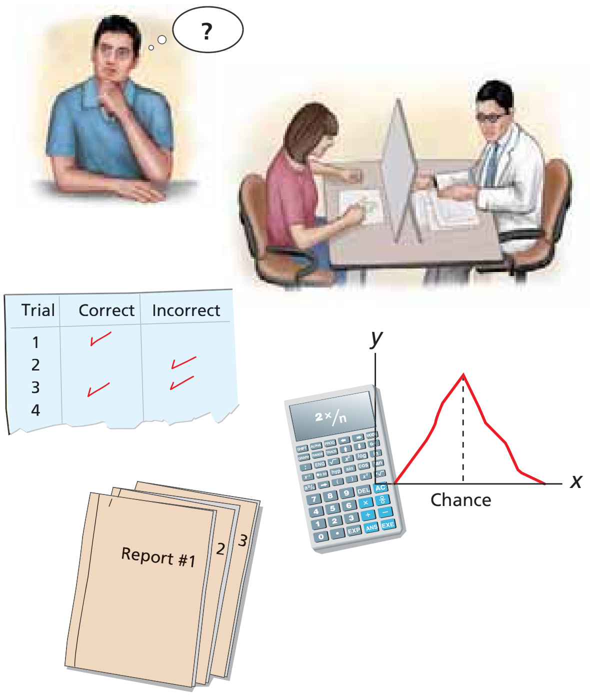


FIGURE 2.1 Five Steps of the Scientific Method

theoretical puzzle. For example, a hypothesis stating that introverted people are attracted to extraverted people might be part of a larger, more complex theory tying together all the factors that affect romantic attraction. Sometimes, however, a hypothesis can be just an interesting idea that piques the scientist's curiosity—as was the case in Emily Rosa's experiment. Her hypothesis came simply from questioning the value of a treatment (therapeutic touch) that everyone "knew" to be effective.

Like any good scientist, Emily stated her hypothesis in such a way that it could be tested and *falsified* (shown to be either correct or incorrect). To make her suspicion testable, Rosa had to follow an ironclad requirement of all scientific research: She had to give **operational definitions** for all the terms in her hypothesis. That is, she had to specify the exact procedures (operations) she would use in setting up the experimental conditions and measuring the results.

Emily wondered: Could TT practitioners accurately sense the presence of her hand when it was placed above one of their hands but out of sight? She hypothesized that they could not. In our earlier example, the study of Clever Hans, Dr. Pfungst also operationalized his hypothesis by stating that the horse could not give the right number of taps with its hoof when it couldn't see its owner or when the owner couldn't see the written questions. Here again, the hypothesis was stated *operationally*—in terms of the procedures that would be used to test it.

So far, so good. But, of course, a scientific study must not stop with a hypothesis. The great failing of pseudosciences like astrology is that they never take the other steps necessary to verify or reject their assertions. Among scientists, however, a hypothesis will be taken seriously only after it has been subjected to rigorous testing.

Performing a Controlled Test A hypothesis must undergo an "ordeal of proof"—a test that it will either pass or fail. Here's how Emily Rosa conducted her test: She invited each of 21 TT practitioners (varying in experience from 1 to 27 years) to determine which of their two hands (thrust, palms up, through holes in a screen) was closest to one of her own hands (held palm down, a few inches from either of the practitioner's hands).

In order to control the conditions of her experiment, Rosa varied only one part of the situation on each trial: whether her hand was above the subject's left or right hand. We call this variable condition the **independent variable (IV).** Think of the independent variable as a condition that the experimenter changes *independently* of all the other carefully controlled experimental conditions. The independent variable always involves a systematic variation on the conditions that the experimenter is evaluating in a study. In Pfungst's study of Clever Hans, the independent variable involved systematically changing the conditions so that (a) Hans could not see his owner or (b) the owner could not see the questions being asked.

In Rosa's experiment on therapeutic touch, control over the experimental conditions would have been laughable if she had simply held her hand alternately above the volunteers' left and right hands or followed some other predictable pattern. That is, had the volunteers been able to guess which response was correct, the results of the experiment would have meant nothing. The solution was **random presentation** of the stimulus, which meant that chance alone determined the order in which the stimulus was presented. Random presentation is one tool in the experimenter's bag of tricks for controlling expectations that can skew the results of a study. In Rosa's experiment, randomization was achieved by a coin flip, which determined whether she presented her hand above the practitioner's left or right hand. And in Pfungst's study, randomization meant that there was no predictable pattern (such as 2, 4, 6, 8 . . .) in the correct answers to the problems presented to Hans and his trainer.

■ **Operational definitions** Specific descriptions of concepts involving the conditions of a scientific study. Operational definitions are stated in terms of how the concepts are to be measured or what operations are being employed to produce them.

■ **Independent variable (IV)** A stimulus condition so named because the experimenter changes it independently of all the other carefully controlled experimental conditions.

■ **Random presentation** A process by which chance alone determines the order in which the stimulus is presented.

Gathering Objective Data In the third step of the scientific method, the scientist collects objective **data**: information gathered by direct observation. Such data depend only on the manipulations of the experimental conditions (the independent variable). The data must *not* depend on the experimenter's hopes, expectations, or personal impressions. In Emily Rosa's experiment, the data consisted of the number of correct and incorrect responses during the test—whether the practitioners responded correctly to the placement of her hand. Such responses are referred to as the **dependent variable (DV).** The term comes from the assumption that the responses of participants in an experiment *depend* directly on the conditions to which they have been exposed. As a result, the data will depend on how the independent variable has been manipulated. (You might think of the independent variable as the *stimuli* you are studying and of the dependent variable as the *responses* made by the participants in your experiment.)

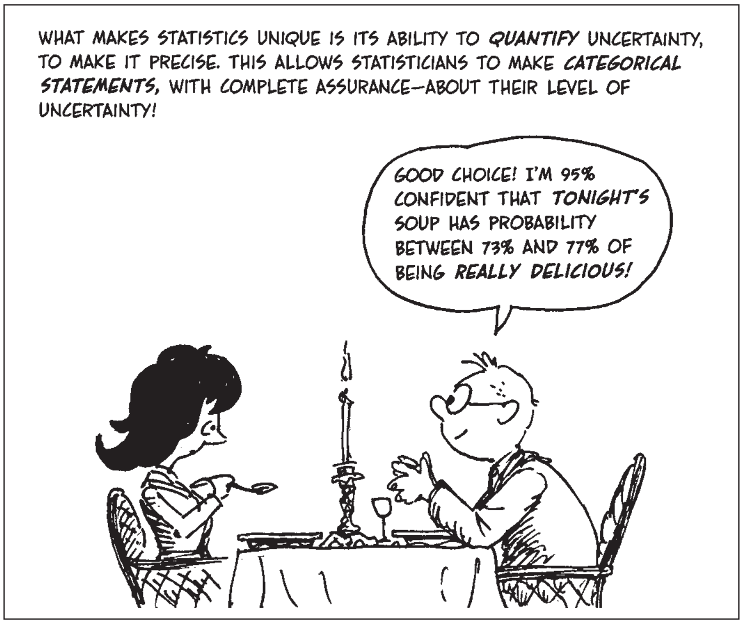
In designing an experiment, the dependent variable must also be given an operational definition. That is, the researcher must specify the procedures (operations) that were used in measuring the responses being observed. This is exactly what Emily Rosa did when she described how she required her participants to respond with guesses of "left" or "right." The dependent variable in Pfungst's study consisted of the horse's hoof-tapping response to each question presented.

Analyzing the Results and Accepting or Rejecting the Hypothesis In the fourth step of the scientific method, the researcher examines the results (the data) to see whether the hypothesis survived the test. Based on that analysis, the hypothesis is accepted or rejected. Making this determination usually necessitates some special mathematical tools, particularly if the data require a close call. Statistical analysis can tell the researcher whether the observed results rise to the level of *significance*—that is, whether the results are likely due to the independent variable or merely due to chance.

A detailed explanation of statistics is beyond the scope of this book. In fact, it's a subject for a whole course in itself. But to give you a glimpse of this world, the second part of this chapter offers a brief introduction to statistics.

■ **Data** Pieces of information, especially information gathered by a researcher to be used in testing a hypothesis. (Singular: *datum.*)

■ **Dependent variable (DV)** The measured outcome of a study; the responses of the subjects in a study.



from "The Cartoon Guide to Statistics" by Larry Gonick & Wollcott Smith

There you will find a summary of key points and examples of how psychological concepts are *quantified* (measured and expressed as numbers) and how those quantities can provide meaning and understanding.

In Rosa's experiment, the statistical analysis was remarkably simple. The chances of getting a correct answer merely by guessing were 50%. That is, half the time the TT practitioners could be expected to give the right answer, even if they had no ability to sense the "human energy field." Accordingly, Rosa set this standard: Her subjects would have to perform significantly above the chance level to support the claim that they can detect a "human energy field." They did not, so she concluded that practitioners of therapeutic touch were not sensing human energy fields.

Much the same analysis applied to Pfungst's study, where the chance level of correct responses would be near zero, and *any* consistent level of correct responses would have supported the hypothesis that Clever Hans could read and calculate. That hypothesis, however, was rejected, because Hans's responses were incorrect when cues from his owner were controlled.

Publishing, Criticizing, and Replicating the Results In the fifth step of the scientific method, researchers must find out whether their work can withstand the scrutiny and criticism of the scientific community. To do so, they might communicate their results to colleagues by publishing them in a professional journal, presenting a paper at a professional meeting, or writing a book. (You may recall that Emily Rosa published her results in the *Journal of the American Medical Association.*) Then they wait for the critics to respond.

If colleagues find the study interesting and important—and especially if it challenges a widely held theory—they may look for flaws in the research design: Did the experimenter choose the participants properly? Were the statistical analyses done correctly? Could other factors account for the results?

Some critics complained that Rosa's experiment was not an accurate representation of the conditions under which therapeutic touch is done: They claimed that TT depends on the transfer of emotional energy during a medical crisis, and because Emily was not sick she didn't have disturbances in her energy field that could be detected by TT practitioners.

Critics could have checked Rosa's work by *replicating* it. To **replicate** her experiment they would redo it, perhaps under slightly different control conditions, to see whether they would get the same results. But as far as we know, Rosa's experiment was never replicated. (Nor was Pfungst's.) At this point, then, we can say that Rosa's experimental results have withstood the scientific test. We should also note that Emily's research earned her a check for $1000 from the Skeptics Society. She also received a plaque from the *Guinness Book of Records* for being the youngest researcher to have a paper published in a major medical journal.

Criticism and replication of research are a part of a thorough, and sometimes intimidating, screening process that goes on behind the scientific scenes to filter out poorly conceived and executed research. As a result, fewer than 2% of the papers submitted to psychological journals get into print without major revisions. In fact, the majority never see print at all (Eichorn & VandenBos, 1985). Journal editors and book publishers (including the publishers of this book) routinely seek the opinion of several expert reviewers for each submission before agreeing to publish it. Different reviewers often focus their criticism on different facets of the study (Fiske & Fogg, 1990). As a result, the author usually receives helpful, if sometimes painful, suggestions for revision. Only when a hypothesis has survived all these tests will editors put it in print and scholars tentatively accept it as scientific "truth." We should emphasize, however, that scientific findings are always tentative—forever in

■ **Replicate** In research, this refers to doing a study over to see whether the same results are obtained. As a control for bias, replication is often done by someone other than the researcher who performed the original study.

jeopardy from a new study that might require a new interpretation and relegate previous work to the scientific scrap heap. Granted, it is an imperfect system, but it is the best method ever developed for testing ideas about the natural world.

Types of Psychological Research

Everything we have ever learned is because of an experiment. Whether we conducted it ourselves or learned it from someone else, it came from an experiment. Some experiments are intentional, such as Newton's experiment to determine the force of gravity or Franklin's discovery of electricity, and some are accidental, such as the discovery of penicillin or how we learned that the stove was hot.

Clearly, research is an essential component of our everyday lives. To be fair, however, not all research is created equal. We look at two methodologies: the experiment, and a variety of quasi-experimental methods. The experiment is probably what first comes to mind when we think of research methodology.

Experimental Method There are many ways to approach the experimental method. Perhaps the best way is to examine the steps in designing an experiment. Table 2.1 lists the components of the research process for an experiment.

The first step in the research process begins with basic inquiry—that is, getting a research idea. What makes you curious? Is there a particular phenomenon that you wonder about, such as learning or remembering? Developing a research question involves generating an hypothesis that is testable, verifiable, and refutable. To determine these things you have to read the literature on your research idea. Some of it can be hard to find, and sometimes it is necessary to narrow down your topic.

After you have developed your hypothesis, you will need to establish your variables. In an experiment we look at three types of variables: independent, dependent, and extraneous (confounding). The independent variable (IV) is the one that the experimenter controls. For example, if you were to conduct an experiment on the effect of light on plant growth, the amount of light provided would be the IV. The dependent variable (DV) is what we measure. In that plant experiment, the amount of growth of the plant is the DV. **Confounding** or **extraneous variables** are other things that can affect the outcome of the experiment. In our example, you would have to ensure that no extra light reached your plants (other than what you specified in your design).

The next task is to ensure that you have **controls**—that is, to ensure that all groups in the experiment are treated exactly the same, except for the IV. In the plant experiment, you would have to ensure that all the plants received the same amount of water, were the same species and age, were exposed to the same temperature, and so on. All of these precautions are necessary so that we can be certain that the data we get at the end of the experiment can be replicated and that our conclusions are valid.

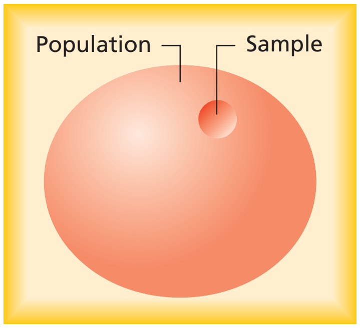
Following the development of your procedures, variables, and controls, you will need to select your subjects. Subjects are drawn from a population, which consists of everyone who fits the description of your test group. For example, if we wanted to test how high school students learn, our population would be all high school students. It would be impossible to test every single high school student, so to compensate for that, we would take a representative sample of the population. (See Figure 2.2.)

■ **Experiment** A kind of research in which the researcher controls all the conditions and directly manipulates the conditions, including the independent variable.

■ **Confounding** or **extraneous variables** Variables that have an unwanted influence on the outcome of an experiment.

■ **Controls** Constraints that the experimenter places on the experiment to ensure that each subject has the exact same conditions.

TABLE 2.1	Components of the Research Process
Developing a research question	
Surveying the literature	
Hypothesis	
Independent variable (IV)	
Dependent variable (DV)	
Extraneous variables	
Controls	
Sampling/Subjects (random assignment to groups)	
Procedure	
Results/Statistics	
Discussion	



● **FIGURE 2.2** A representative sample consists of randomly selected individuals who accurately represent the targeted population.

To do that we would randomly select individuals who accurately represent the population of high school students. This is a time-consuming and costly process, but it ensures that that our data will reflect the results we would get if we tested everyone. The ability to choose anyone from the population is referred to as random selection.

Another key thing to keep in mind is that every subject of an experiment must have an equal chance of being in the experimental group (which receives the IV) or the control group (which receives either a placebo or nothing). This is called **random assignment.** Each member of the sample has an equal likelihood of being chosen for the experimental group.

Once the subjects have been selected and assigned, the experiment can begin. Carefully following the procedures laid out in the research design, experimenters conduct their experiment. In their design, they must account for, and try to control, as many aspects of the experiment as possible. The reason for this is to try to remove the influence of outside influences on the experiment. As we have seen, these influences are called *extraneous variables*. The use of controls limits the influence of these so that the results of the experiment accurately reflect what is being tested.

At the conclusion of the experiment, data is collected and subjected to statistical analysis.

Non-Experimental Methods Given that experiments—painstakingly applied—can yield true cause-and-effect statements about a situation, why wouldn't we always do experiments? There are times when we cannot do so for ethical or practical reasons. Take, for example, research on cancer. To conduct a true experiment on a possible cure for lung cancer, we would have to give people lung cancer and then try to cure it. This, of course, would be morally indefensible. Instead, we would choose subjects based on the preexisting condition of having lung cancer, and try to cure them. This design is a *non-experimental* design.

Non-experimental methods can yield useful data. But, they are just not true experiments because they are missing a component of the experiment, usually in the area of controls. Of the seven non-experimental methods listed in Table 2.2, none should be disregarded; rather, all of them need to be applied carefully and with a great deal of planning.

Take the cancer experiment mentioned above. This design, in which we choose subjects on the basis of a preexisting condition, is called **ex post facto** research. We chose this method mainly in response to ethical considerations. Because our treatment for lung cancer is not guaranteed (it is being tested), the treatment might not work. In addition, ethical considerations arise in *not* giving someone the best possible treatment!

■ **Random assignment** Each subject of the sample has an equal likelihood of being chosen for the experimental group of an experiment.

■ **Ex post facto** Research in which we choose subjects based on a pre-existing condition.

■ **Correlational study** A type of research that is mainly statistical in nature. Correlational studies determine the relationship (or correlation) between two variables.

TABLE 2.2 Seven Non-Experimental Methodoligies

- Ex-post facto design
- Correlation study
- Survey
- Naturalistic observation
- Longitudinal study
- Cross-sectional study
- Cohort-sequential study

Correlational Studies An alternative "quasi" design is called a **correlational study.** What we are doing here is seeing the relationship (or correlation) between two variables. For example, when the surgeon general first began telling people that smoking and lung cancer were related, that statement was based on the correlation between people who smoke and the incidence of lung cancer among those people. Not everyone who smokes gets lung cancer, and not everyone who gets lung cancer smokes, but there is a very high correlation between smoking and lung cancer. As scientists often put it, *correlation does not necessarily mean causation.*

Scientists usually express the degree of correlation as a number. This requires calculating a statistic known as the *correlation coefficient*, often symbolized in formulas as the letter *r*. The correlation coefficient summarizes the relationship between the two variables. It can range from a negative number as low as –1.0 to a positive number as high as +1.0.

We won't go into the details of calculating the correlation coefficient here. The important idea is to develop a feeling for what *positive correlation*, *negative correlation*, and *zero correlation* mean. If the variables have no relationship at all, their correlation is 0. You would expect a zero correlation between shoe size and GPA, for example. If, however, the two variables show a relationship in which they vary in the same direction (as the values of one variable increase, so do those of the other), then we say they have a positive correlation. An example of a positive correlation is the moderate relationship (approximately +0.4) between SAT scores and college grades.

It is important to understand that *a correlation can show a strong relationship even when it is negative.* Let us suppose that a measure of anxiety (such as a checklist of anxiety-related symptoms) shows a correlation of −0.7 between anxiety and time spent studying. In other words, more study is associated with less anxiety. Even though this is a negative correlation, it shows a *stronger* relationship than, for example, the positive correlation between SAT scores and grades (+0.4).

Another research method is the **survey.** Again, a survey is not a true experiment, but if conducted correctly, it can yield useful data. When designing a survey, the researcher must take great care to make sure the questions are not skewed or biased toward a particular answer. Also, when comparing survey results, one needs to go back and look at both surveys to make sure that the questions and answer scales are parallel. You need to be certain that questions were being asked the same way in order to be able to compare the results and draw conclusions. We will give a detailed example of the survey method later in the chapter.

In **naturalistic observation,** subjects are observed in their natural environment. This method is a good choice for studying, say, child-rearing practices, people's shopping habits, or public courting behaviors. In order to remove demand characteristics, which are cues the experimenter inadvertently gives that tell the subject what "good" results are, the subjects in a naturalistic observation should not know they are being observed. This ensures that the behavior being observed is the actual behavior in its natural state. This approach is also used extensively to study animal behavior in the wild. (Jane Goodall used it in her classic studies of chimpanzee culture.) Because a researcher merely observes, rather than controlling the conditions or manipulating the independent variable, naturalistic observations are made under far less controlled conditions than experiments.

What if you wanted to investigate the long-range effects of something? The type of research you might be most interested in is called a **longitudinal study.** In this type of study, one group of subjects is followed and observed (or examined, surveyed, etc.) for an extended period of time, such as 20 years. The benefit of this research is that you have the same subject group throughout. The drawbacks are time and expense.

The next two methodologies were developed to avoid the time and expense of the longitudinal study. The **cross-sectional study** examines a representative cross section of the population and tests/surveys these subjects at one specific time. This will yield data similar to longitudinal data but not so accurate. The **cohort-sequential study** yields better data. Here the investigators take a cross section of the population and then follow each *cohort* or group for a short

■ **Survey** A quasi-experimental method in which questions are asked to subjects. When designing a survey, the researcher has to be careful that the questions are not skewed or biased toward a particular answer.

■ **Naturalistic observation** A research method in which subjects are observed in their natural environment.

■ **Longitudinal study** A type of study in which one group of subjects is followed and observed (or examined, surveyed, etc.) for an extended period of time (years).

■ **Cross-sectional study** A study in which a representative cross section of the population is tested or surveyed at one specific time.

■ **Cohort-sequential study** A research method in which a cross section of the population is chosen and then each cohort is followed for a short period of time.

Jane Goodall used the method of naturalistic observation to study chimpanzee behavior.

period of time. This study can take less time than the longitudinal design, is much less susceptible to bias, and therefore yields more accurate data than a cross-sectional study.

Sources of Bias in Research (or Anywhere Else)

Think of an issue on which you have strong feelings and opinions—perhaps abortion, euthanasia, or capital punishment. On such topics, our emotions make it difficult to reason objectively. Likewise, emotionally loaded topics can bring out biases that affect the ways an experimenter designs a study, collects the data, or interprets the results. Fortunately, the scientific method, with its public procedures and openness to replication, provides a powerful means to check on an experimenter's bias. Still, scientists would rather save themselves embarrassment by identifying and controlling their biases before they hit print. Here are some forms of bias to which they must be alert.

Personal bias involves an individual's beliefs, preferences, assumptions, or prejudices. Often these are not obvious to the individual holding such biases. For example, in his book *Even the Rat Was White,* psychologist Robert Guthrie (1998) points out the personal bias in the long tradition of using mainly white subjects in psychological research. Whatever form it takes, personal bias can cause scientists to notice only the evidence confirming their hypotheses and to ignore contrary data.

Expectancy bias also affects observations when observers expect—and look for—certain outcomes. We can see expectancy bias at work in a classic study in which psychology students timed groups of rats running through a maze (Rosenthal & Jacobson, 1968a). The experimenters told some students that their rats were especially bright; other students heard that their rats were slow learners. (In fact, the experimenters had randomly selected both groups of rats from the same litters.) Amazingly, the students' data showed that rats believed to be bright outperformed their supposedly duller littermates.

These sources of bias not only can lead to erroneous conclusions but also can be expensive. Imagine that you are a psychologist working for a pharmaceutical company that wants you to design a test for a new drug. With millions of dollars riding on the outcome, you will want to do it right. But what about the doctors who are going to be prescribing the drug to patients in your study? Surely those doctors will have high hopes for the drug, as will their patients. And so the stage is set for bias to creep into your study along with people's expectations.

We have seen that a common strategy for controlling expectancy bias in a drug study is to keep participants in the research experimentally "blind," or uninformed, about whether they are getting the real drug or a placebo. An even better strategy is to keep *both* the participants and the experimenter clueless about which group receives what treatment. In a drug study, this would mean that neither the researchers nor the participants would know (until the end of the study) which individuals were getting the new drug and which were getting the placebo. Such a research strategy is called a **double-blind study.** This strategy ensures that the experimenters will not inadvertently treat the experimental group differently from the control group, so that neither group will have any idea about the expected response to the pills they are taking.

Aside from these forms of observer bias, researchers must also try to identify other possible influences on the behavior being studied—influences other than the independent variable. Such *confounding variables* are factors that could be confused with the independent variable and thus distort the results. Consider, for example, a study of a stimulant drug (such as Ritalin) used to control hyperactive behavior among schoolchildren. What might be some confound-

■ **Personal bias** The researcher allowing personal beliefs to affect the outcome of a study.

■ **Expectancy bias** The researcher allowing his or her expectations to affect the outcome of a study.

■ **Double-blind study** An experimental procedure in which both researchers and participants are uninformed about the nature of the independent variable being administered.

ing variables? The drug's effect might differ because of different body weights, eating schedules, or time, method, or setting of administration. Unless arrangements are made to control all such possible confounding variables—that is, to expose all the subjects to identical conditions—the researcher has no way of knowing which factors really produced the results.

CONNECTION: CHAPTERS 5 AND 13

Paradoxically, stimulants seem to calm hyperactive behavior in children with ADHD.

Ethics in Research

Ethical considerations are an overarching component of all research. These issues range from the basic question "Should the research be conducted?" (as is being argued today regarding the use of stem cells) to questions such as "Should research be approved even if there is no direct application for it?" (in essence, the issue of basic versus applied research). Dess and Foltin (2005) pose seven questions involving what they call the "Ethics Cascade" (see Table 2.3).

The questions posed here are not simple, nor are their answers. Ethical guidelines such as the APA's "Ethical Principles of Psychologists and Code of Conduct" (2002) must be followed in the conduct of *all* research. Each research institution must have an **Institutional Review Board (IRB)** that reviews and approves all research. In addition, animal research must also be approved by an **Institutional Animal Care and Use Committee (IACUC).** Gruber (2005) also points out that all animal research must comply with the *ABCs of laboratory animal research* (Appropriate, Beneficial, and Caring).

IACUCs and IRBs are put in place to ensure not only that researchers and institutions comply with federal, state, and local laws and regulations, but also that all research is conducted ethically and humanely. No researcher takes his or her work lightly, and the ethics involving all research—be it human or animal—is serious indeed.

Deception The use of *deception* poses an especially knotty problem. The APA's "Ethical Principles" states that under most circumstances, participation in research should be voluntary and informed. That is, we should advise volunteers of what challenges they will face and give them a real opportunity to drop out of the study if they want to. But what if you are interested in the "good Samaritan" problem—the conditions under which people will help a stranger in distress? If you tell people that you have contrived a phony emergency situation and ask them whether they are willing to help, you will spoil the very effect you are trying to study. Consequently, the guidelines do allow for deception under some conditions, provided that no substantial risks are likely to accrue to the participants.

■ **Institutional Review Board (IRB)** A committee at each institution where research is conducted to review every experiment for ethics and methodology.

■ **Institutional Animal Care and Use Committee (IACUC)** A committee at each institution where research is conducted to review every experiment *involving animals* for ethics and methodology.

TABLE 2.3 Summary of Steps in the Ethics Cascade

I. Who should decide what is morally justifiable in the conduct of research?
II. Are controlled research studies ever necessary or appropriate?
III. Should all research have a foreseeable practical benefit?
IV. At whom should research be directed?
V. What specific topics are worthy of research?
VI. What particular research methodologies are scientifically valid, as well as ethically appropriate?
VII. Of the valid methods, which should be used?

Source: From LABORATORY ANIMALS IN RESEARCH AND TEACHING, Ethics, Care, and Methods edited by Chana K. Akins, Sangeeta Panicker, Christopher L. Cunningham. Chapter 2, The Ethics Cascade, Nancy K. Dess and Richard W. Foltin. APA Press, Washington, DC, 2004, reprinted with permission.

People are sharply divided on the use of laboratory animals in research.

When deception is used, the APA guidelines require that participants be informed of the deception as soon as is possible without compromising the study's research goals. Individuals used in deceptive research must also be *debriefed* after the study to make sure that they suffer no lasting ill effects. Despite these precautions, some psychologists remain opposed to the use of deception in any form of psychological research (Baumrind, 1985; Bower, 1998d).

Animal Studies Another long-standing ethical issue surrounds the use of laboratory animals, such as rats, pigeons, and monkeys. As far back as the mid-1800s, scientists used animals in their research for a variety of reasons. These included the relative simplicity of animals' nervous systems and the relative ease with which a large number of individuals could be maintained under controlled conditions. Animals have also served as alternatives to human subjects when a procedure was deemed risky or outright harmful. Concerned about the issue as long ago as 1925, the American Psychological Association established a Committee on Precautions in Animal Experimentation, which adopted guidelines for animal research (Dewsbury, 1990). The American Psychological Association's "Ethical Principles of Psychologists" (2002) directs researchers to provide decent living conditions for animal subjects and to weigh any discomfort caused them against the value of the information sought in the research. A 1985 federal law also imposes legal restrictions on animal research (Novak & Suomi, 1988).

Recent years have seen a renewal of concern, both inside and outside of psychology, about the use of animals as subjects, particularly when the research involves painful or damaging procedures, such as brain surgery, electrode implants, and pain studies. Some people feel that the limitations should be more stringent on studies using humanlike animals, such as chimpanzees. Others believe that limitations or outright bans should apply to all animal research, including studies of simple animals such as sea slugs (which are often used in neurological studies). Many psychologists, however, support animal research under the APA guidelines (Blum, 1994). Heated debate about this issue continues.

TABLE 2.4	What Questions Can the Scientific Method *Not* Answer?

The scientific method is not appropriate for answering questions that cannot be put to an objective, empirical test. Here are some examples of such issues:

Topic	Question
Ethics	What are the ethical issues involved in animal research?
Values	Which culture has the best attitude toward work and leisure?
Morality	When is it morally acceptable to go to war?
Preferences	Is rap music better than blues?
Aesthetics	Was Picasso more creative than Van Gogh?
Existential issues	What is the meaning of life?
Religion	How do people of faith explain natural disasters?
Law	What should be the speed limit on interstate highways?

Although science can help us understand such issues, the answers ultimately must be settled by logic, faith, legislation, consensus, or other means that lie beyond the scope of the scientific method.

Questions Science Cannot Answer

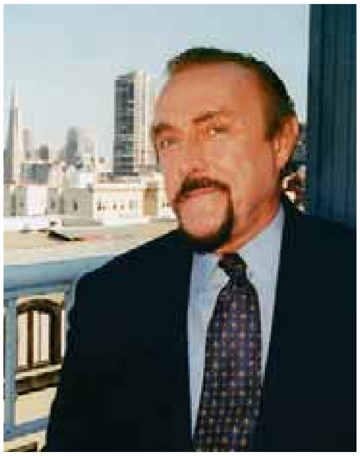
It is important to understand that science is not the best approach for finding answers to every important question in our lives. Even scientists don't take a scientific approach to everything. The scientific method is merely the best way to find answers to testable questions about the natural world—the world of atoms and animals, of stones and stars, and of behavior and mental processes. On the other hand, science is *not* appropriate for answering questions that cannot be empirically tested—such as questions of ethics, morality, religious beliefs, or preferences. To see what we mean, please look at Table 2.4, which shows some of the questions that science can never answer.

PSYCHOLOGY IN YOUR LIFE: GETTING IN DEEPER

Whatever your intended major field of study, you will want to learn more about the professional role your chosen field will expect you to play. You can do this in several ways: by attending events sponsored by your major department, by getting to know your professors personally, and by taking out student memberships in professional organizations. You should also develop a habit of scanning the field's main magazines, journals, and newsletters. For those readers who are considering a major in psychology, we suggest that you investigate the following resources.

Professional Organizations in Psychology The largest and oldest professional association for psychologists, the American Psychological Association (APA), has well over 150,000 members and affiliates (American Psychological Association, 2004). The American Psychological Society (APS) was formed just a few years ago to give a stronger voice to academic and research psychologists. Although the APS is a much smaller organization, it has won wide respect; many psychologists belong both to the APA and to the APS.

These groups have student memberships that include nearly all privileges at a fraction of full membership costs. If you are thinking of majoring in psychology, ask your instructor for information about student membership in a



Dr. Phil Zimbardo, one of your authors, served as president of the American Psychological Association in 2002.

professional psychology association. Also consider attending a state, regional, or national convention to get a better sense of what psychologists are really like. These conventions also offer an opportunity for students to present their own research. You could do so, too.

Consider, also, joining a student psychology group, if your school has one. If none is available, you may be able to organize a psychology club or a chapter of a national honorary society, such as Psi Beta (at a two-year college) or Psi Chi (at a four-year college or university).

Psychology-Related Journals and Magazines Professional groups sponsor newsletters or journals that help keep their members abreast of new developments in the field. Psychology majors should begin looking over a few of the main ones every month. Some publish general-interest articles; others contain highly technical reports tailored for those with specialized advanced training. We suggest taking your first plunge into the psychological literature with one or more of these:

- *Monitor on Psychology*—the monthly news magazine of the APA
- *Current Directions in Psychological Science*—a semimonthly APS journal that provides short reviews on trends and controversies in all areas of psychology
- *American Psychologist*—the flagship journal of the APA
- *Psychological Science*—the premiere journal of the APS
- *Whitman Journal of Psychology*—a biannual journal of high school research

In addition, there are several popular magazines in which you may find psychological articles of interest:

- *Discover*—a science magazine written for the general public
- *Scientific American*—another general-interest science magazine
- *Science News*—a weekly magazine consisting of brief blurbs on breaking news in all areas of science, including psychology
- *The Skeptical Inquirer*—a take-no-prisoners, pseudoscience-bashing magazine published by CSICOP, the Committee for the Scientific Investigation of Claims of the Paranormal

Don't feel that you must keep up on the entire psychological literature. Nobody can. Read what interests you in these publications.

Electronic Resources in Psychology The printed psychological literature is vast and growing quickly. As a result, anyone wanting to find out what is known on a special topic must know how to access the information on the Internet and in an electronic database. There are several general databases available, such as Expanded Academic Index and Ebsco Academic Search Elite. The best electronic resource specifically for psychology is PsychInfo, an online computer database offered by the American Psychological Association. Most such resources require a paid subscription, although they may be available through your campus library.

In addition, a huge amount of free information about psychology is available on the Internet. A good place to start looking would be the American Psychological Association's home page on the World Wide Web at http://www.apa.org or the American Psychological Society's home page at http://www.psychologicalscience.org. Remember that Web addresses often change. Remember, also, that anyone can put anything on the Internet, so be skeptical!

CHECK YOUR UNDERSTANDING

1. **RECALL:** A theory is
   - **a.** an unsupported opinion.
   - **b.** a testable explanation for what has been observed.
   - **c.** the opposite of a fact.
   - **d.** a statement that has not yet been supported with facts.
   - **e.** an experimental supposition.
2. **RECALL:** A scientific study should begin with
   - **a.** a controlled test.
   - **b.** a hypothesis.
   - **c.** data collection.
   - **d.** risk/gain assessment.
   - **e.** background reading.
3. **APPLICATION:** Which of the following could be an operational definition of "fear"?
   - **a.** an intense feeling of terror and dread when thinking about some threatening situation
   - **b.** panic
   - **c.** a desire to avoid something
   - **d.** moving away from a stimulus
   - **e.** moving toward a stimulus
4. **ANALYSIS:** The conditions involving the independent variable could also be thought of as
   - **a.** cognitions.
   - **b.** experimenter biases.
   - **c.** responses.
   - **d.** results.
   - **e.** stimuli.
5. **RECALL:** Which is the only form of research that can determine cause and effect?
   - **a.** a case study
   - **b.** a correlational study
   - **c.** an experimental study
   - **d.** a naturalistic observation
   - **e.** a survey
6. **ANALYSIS:** Random assignment of subjects to different experimental conditions is a method for controlling differences between
   - **a.** the dependent variable and the independent variable.
   - **b.** the experimental group and the control group.
   - **c.** empirical data and subjective data.
   - **d.** heredity and environment.
   - **e.** controls and extraneous variables.
7. **RECALL:** In which kind of research does the scientist have the most control over variables that might affect the outcome of the study?
   - **a.** a case study
   - **b.** cohort-sequential study
   - **c.** a correlational study
   - **d.** an experimental study
   - **e.** a naturalistic observation
8. **ANALYSIS:** Which one of the following correlations shows the strongest relationship between two variables?
   - **a.** +0.4
   - **b.** +0.38
   - **c.** −0.7
   - **d.** 0.05
   - **e.** −0.9
9. **ANALYSIS:** Which one of the following is a good method for controlling expectancy bias?
   - **a.** performing a case study
   - **b.** joining a professional organization
   - **c.** consulting the APA's "Ethical Principles of Psychologists and Code of Conduct"
   - **d.** doing a double-blind study
   - **e.** clearly describing the intended results to the subjects

ANSWERS: 1.b 2.b 3.d 4.e 5.c 6.b 7.d 8.e 9.d

HOW DO WE MAKE SENSE OF THE DATA?

A longitudinal study was conducted at Bennington College examining political views of students and how they are influenced by campus culture. This longitudinal study was begun in the 1930's, and continued through 1984. The authors of the study, Alwin, Newcomb, and Cohen, found that students' political views can be profoundly influenced by their campus culture—which should make it interesting for you to think about the climate of political opinion of your current school, and of your future college. Do the students and faculty at your school lean toward the liberal or the conservative end of the spectrum? And are the students at your school typical of their counterparts elsewhere in the country? In the following pages we will use these questions as a starting point for an exploration of the statistical methods psychologists use to make sense of the data they gather in their research.

Every fall, the *Chronicle of Higher Education* publishes its "Almanac Issue," which reports the results of a survey of first-year students at colleges and universities across the country. Table 2.5 shows how a national sample stood on a number of political issues (*Chronicle of Higher Education*, 2004). We will use this survey as the basis for assessing the political attitudes of your classmates and comparing them with those of other students across the United States.

We will begin by converting the items in the national survey into a scale that measures liberal and conservative attitudes. The second step will be to determine how you might use that scale to assess your psychology class or some other sample of students at your college or university. Next, we will show you how the resulting data might be organized and analyzed so that you could compare your own survey results with the national student survey data. In addition, we will discuss how your data could be linked, or *correlated*, with other measures, such as income, gender, or grade-point average. Then, in the final part of this section, we will point out some of the statistical pitfalls into which the unwary researcher may fall.

CORE CONCEPT

Researchers use statistics for two major purposes: (1) descriptively to characterize measurements made on groups or individuals and (2) inferentially to judge whether those measurements are the result of chance.

Statistics can be used in a myriad of ways. Perhaps the most obvious one is through the use of surveys.

Developing Your Own Survey

A look at Table 2.5 reveals that the items on the national survey are written in two different ways. Some questions are worded so that agreement is a "conservative" response. (Item 1 is an example of conservative wording: "There is

TABLE 2.5 National Student Survey Data

Agree strongly or somewhat that:	Conservative/ liberal wording	Agree	Majority response	Liberal response[a]
1. There is too much concern in the courts for the rights of criminals.	Conservative	61.1%	Conservative	38.9%
2. Abortion should be legal.	Liberal	54.5%	Liberal	54.5%
3. The death penalty should be abolished.	Liberal	32.6%	Conservative	32.6%
4. Marijuana should be legalized.	Liberal	38.8%	Conservative	38.8%
5. It is important to have laws prohibiting homosexual relationships.	Conservative	26.1%	Liberal	73.9%
6. The federal government should do more to control the sale of handguns.	Liberal	76.5%	Liberal	76.5%
7. Racial discrimination is no longer a problem in America.	Conservative	22.4%	Liberal	77.6%
8. Wealthy people should pay a larger share of taxes than they do now.	Liberal	53.1%	Liberal	53.1%
9. Same-sex couples should have the right to legal marital status.	Liberal	59.4%	Liberal	59.4%
10. Affirmative action in college admissions should be abolished.	Conservative	52.8%	Conservative	52.8%
11. The activities of married women are best confined to the home and family.	Conservative	21.7 %	Liberal	78.3%
12. Federal military spending should be increased.	Conservative	38.4%	Liberal	61.6%

[a]The score in this column has been converted to the same terms as the Liberal–Conservative Scale (LCS) scoring system. So when the question has been worded conservatively, the LCS score is calculated by subtracting the percentage who agree from 100.

too much concern in the courts for the rights of criminals.") Other items are worded so that agreement is a "liberal" response. (Item 3, for example, is worded in the liberal direction: "The death penalty should be abolished.") Good surveys are constructed in this way to be neutral and to control for the tendency some people have of simply agreeing or disagreeing with each statement.

In Table 2.5 we have indicated on each item whether agreement with the statement indicates a liberal or a conservative attitude. While you may disagree with our judgment about the liberalness or conservativeness of a particular item, it is important to note that you can clearly see what we mean by "liberal" and "conservative" by the way we have designated each item. By doing so, we have given *operational definitions* of the terms *liberal* and *conservative.* Together these items comprise what we will call our Liberal–Conservative Scale (LCS). By administering the LCS to your class, you can not only obtain political attitude scores for students in your class but also compare the class's responses with the national survey data.

To score the responses obtained on the LCS, we will give one point for each of the following "liberal" items with which a respondent *agrees:* 2, 3, 4, 6, 8, and 9. Further, we will give another point for each of the following "conservative" items with which a respondent *disagrees:* 1, 5, 7, 10, 11, and 12. Accordingly, high scores will indicate a liberal tendency, and low scores will indicate a conservative tendency. (There is no value judgment here: Neither a high nor a low score is judged as being better.)

To illustrate how we might use the LCS in a study of students' political attitudes, let's suppose that we have administered the LCS to a class of 50 students. The resulting data (which we have contrived) appear in Table 2.6. By merely counting the questions on which the majority gave a liberal response, we find that our class was more conservative (with conservative majorities on six items) than the national sample (which had liberal majorities on eight items).

Although this is an interesting result, there is much more that can be learned by organizing the data obtained from our survey. Let's first take a look, in the next section, at the *raw data*.

Organizing the Data

In addition to the data showing how students responded on each question, we obtained the following set of LCS scores for the class:

4	4	8	3
3	3	7	9
6	7	8	9
6	6	1	4
3	6	11	8
7	1	8	5
8	7	5	7
1	1	6	5
4	6	10	7
9	8	5	8
2	4	8	4
5	6	2	3
10	2		

TABLE 2.6 Distribution of LCS Responses from a Class of 50 College Students

Agree strongly or somewhat that:	Number of respondents who agree	Agree	Majority response	Liberal response
1. There is too much concern in the courts for the rights of criminals.	36	72%	Conservative	28%
2. Abortion should be legal.	28	56%	Liberal	56%
3. The death penalty should be abolished.	19	38%	Conservative	38%
4. Marijuana should be legalized.	24	48%	Conservative	48%
5. It is important to have laws prohibiting homosexual relationships.	16	32%	Liberal	68%
6. The federal government should do more to control the sale of handguns.	24	48%	Conservative	48%
7. Racial discrimination is no longer a problem in America.	14	28%	Liberal	72%
8. Wealthy people should pay a larger share of taxes than they do now.	24	48%	Conservative	48%
9. Same-sex couples should have the right to legal marital status.	32	64%	Liberal	64%
10. Affirmative action in college admissions should be abolished.	31	62%	Conservative	38%
11. The activities of married women are best confined to the home and family.	13	26%	Liberal	74%
12. Federal military spending should be increased.	19	38%	Liberal	62%

As you can see immediately, a set of raw data in this form is nearly impossible to interpret. Accordingly, our first task is to arrange the LCS scores into a **frequency distribution,** as shown in Table 2.7. In the "Frequency" column of the table, you will see, for example, that four students received a score of 1, three scored 2, and so on. Grouping of the data in this way makes much more sense than did the array of raw data above. Going one step further, we can convert the data into a bar graph called a **histogram,** which we have drawn in Figure 2.3. In this diagram, you can more readily see that the students' scores are not evenly distributed across the scale. The histogram also makes it obvious that the scores are more clustered near the middle of the distribution than they are at the ends.

■ **Frequency distribution** A summary chart, showing how frequently each of the various scores in a set of data occurs.
■ **Histogram** A bar graph depicting a frequency distribution. The height of the bars indicates the frequency of a group of scores.
■ **Descriptive statistics** Statistical procedures used to describe characteristics and responses of groups of subjects.

Describing the Data with Descriptive Statistics

We can bring our data into even sharper focus by calculating some simple **descriptive statistics,** which are numbers that describe the main characteristics of the data. In particular, psychologists often find it useful to find a number that represents the middle of a distribution—the central point around which the scores seem to cluster. This is called a *measure of central tendency.* Additionally, researchers usually want a statistic that indicates the spread of the dis-

TABLE 2.7	Frequency Distribution of LCS Scores for a Class	
LCS score	Frequency	LCS score × frequency
1	4	4
2	3	6
3	5	15
4	6	24
5	5	25
6	7	42
7	6	42
8	8	64
9	3	27
10	2	20
11	1	11
12	0	0
	$\Sigma = 50$	$\Sigma = 280$

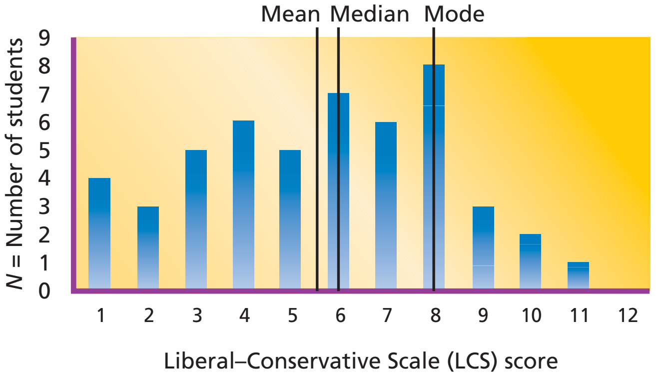


FIGURE 2.3 Distribution of Liberal–Conservative Scale (LCS) Scores for a Hypothetical Class

The height of each bar indicates how many respondents obtained exactly that score on the LCS. Note that the three averages, the *mean,* the *median,* and the *mode,* are in different locations with the distribution not perfectly symmetrical. The mean is heavily influenced by extreme outlying scores, such as the four students who scored 1 on the LCS.

tribution—how closely the scores bunch up around the central point. This is called a *measure of variability.*

Measures of Central Tendency: Finding the Center of the Distribution You are undoubtedly more familiar with the everyday name for measures of central tendency: *averages.* As their more formal name suggests, measures of central tendency help us locate the center of a set of measurements, such as we have from the responses we obtained on the Liberal–Conservative Scale. Three forms of central tendency are most commonly used: the *mean,* the *median,* and the *mode.* Let's look briefly at each in turn.

The Mean Most people think only of the **mean** when they hear the word *average.* The mean is, no doubt, familiar to you as the statistic used to calculate your grade-point average. And it is the statistic that psychologists most often use to describe sets of data. To find the mean, you simply add up all the scores in a distribution and divide by the total number of scores. The calculation is summarized by the following formula:

$$M = \Sigma X \div N$$

Here M is the mean, Σ (the Greek capital letter *sigma*) is the summation of what immediately follows it, X represents each individual score, and N represents the total number of scores. In our example, to calculate the mean we would first add up all the Liberal–Conservative Scale scores (ΣX). The resulting sum is 280. Then we would divide that sum by the total number of scores ($N = 50$). Thus our mean (M) of the LCS scores for the class would be

$$M = 280 \div 50 = 5.6$$

Usually the mean is a good indicator of the center of the distribution, as you will note in Figure 2.3. Unfortunately, it has one potential flaw: Under some circumstances the mean can be unduly influenced by extreme scores. When the distribution is relatively symmetrical, this is not a problem. But when the scores bunch up toward one end of a distribution (in a *skewed* distribution),

Mean The measure of central tendency most often used to describe a set of data—calculated by adding all the scores and dividing by the number of scores.

a few extreme scores at the other end can have a disproportionate effect that pulls the mean toward the extreme score. Because of this effect, researchers sometimes choose one of the other measures of central tendency to find an average for a highly skewed distribution.

The Median One of the alternative measures of central tendency is the **median,** the middle score—the score that separates the upper half of the distribution from the lower half. In our example, the median is 6. That is, half of the scores are 6 or higher, and the other half are 6 or lower. (See Figure 2.3.) The big advantage of the median is that it is not distorted by extreme scores.

The Mode The third and simplest of the averages, or measures of central tendency, is called the **mode.** It is merely the score that occurs more often than any other. In our data, more students received a score of 8 than any other number, as shown in Figure 2.3. The modal response for this class on our conservative–liberal scale, therefore, is 8. Although the mode is the easiest index of central tendency to determine, it is often the least useful, especially when the sample is relatively small.

Take a look again at the distribution of scores in Table 2.7 and Figure 2.3. Which of the averages seems to fit the distribution best? Is it the mean of 5.6, the median of 6, or the mode of 8?

Using Averages How can we use averages to compare the class we tested with student responses on the national survey? As shown in the last two columns of Table 2.4, it is easy to convert the national survey percentages to indicate liberal responses—much as we did for the LCS scores. The mean of these national percentages is 58.2. That figure is higher than the mean of our own data set, which is 53.7. These two scores jibe with our earlier comparison of the two groups and further confirm that our class gave, on the average, more conservative responses than the national sample.

Measures of Variability: Finding the Spread of the Distribution In addition to knowing which score best represents the distribution's center, it is often useful to know how well the average represents the distribution as a whole. That is, we may want to know whether most of the scores cluster closely near the average or whether they are spread widely. We use statistics called *measures of variability* to describe the "spread-outness," of scores around some measure of central tendency.

To illustrate why variability is important, suppose that you are a third-grade teacher, and it is the beginning of the school year. Knowing that the average child in your class can read a third-grade-level book will help you to plan your lessons. You could plan more effectively, however, if you knew how similar or how divergent the reading abilities of the 30 children are. Do they all read at about the same level—that is, do they have *low variability?* If so, then you can develop a fairly standard third-grade lesson. But what if the group has *high variability,* with several who can read fourth-grade material and others who can barely read at all? In the latter case, the average reading level is not so representative of the entire class, and you will have to plan a variety of lessons to meet the children's varied needs.

The simplest measure of variability is the **range,** the difference between the highest and the lowest values in a frequency distribution. Returning to the scores produced by our hypothetical class on the Liberal–Conservative Scale, you can see in Figure 2.4 that the scores range from 1 to 11. Thus, to compute the range, you need know only two scores, the highest and the lowest.

While the range is simple to determine, psychologists usually prefer measures of variability that take into account all the scores in a distribution, not just the extremes. The most widely used alternative is the **standard deviation (SD),**

■ **Median** A measure of central tendency for a distribution, represented by the score that separates the upper half of the scores in a distribution from the lower half.

■ **Mode** A measure of central tendency for a distribution, represented by the score that occurs more often than any other.

■ **Range** The simplest measure of variability, represented by the difference between the highest and the lowest values in a frequency distribution.

■ **Standard deviation (SD)** A measure of variability that indicates the average difference between the scores and their mean.

a measure of variability that shows an average difference between each score and the mean. To calculate the standard deviation of a distribution, you need to know the mean of the distribution, along with the individual scores. Although the arithmetic involved in calculating the standard deviation is easy, the formula is a bit more complicated than the one used to calculate the mean and will not be presented here. The general procedure, however, involves subtracting the value of each individual score from the mean and then determining an average of those mean deviations. (Many calculators have a button for computing the standard deviation of a set of scores.)

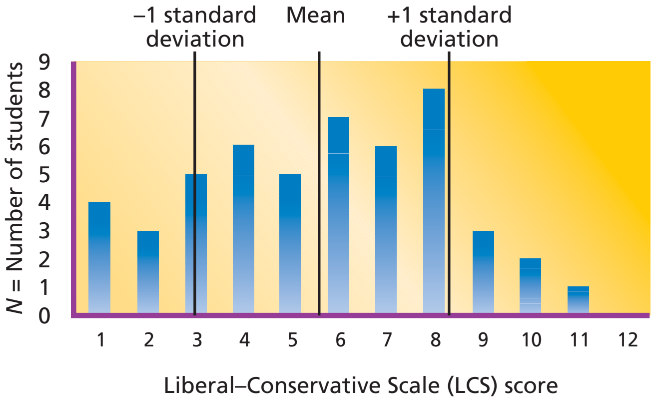


FIGURE 2.4 The Range and Standard Deviation: Distribution of Liberal–Conservative Scale (LCS) Scores for a Hypothetical Class

The range and standard deviation are both measures of the variability or spread of a distribution. Note that most of the scores lie within one standard deviation of the mean.

Happily, the standard deviation is easy to interpret. The larger the standard deviation, the more spread out the scores are; the smaller the standard deviation, the more the scores bunch together around the mean. In our example, the standard deviation of the LCS scores is approximately 2.6. This indicates that approximately two-thirds of the group's scores can be found within 2.6 points of the mean (which is 5.6). To say the same thing in different words, about two-thirds of the scores in our distribution lie between 3 and 8.2.

Together, the mean and the standard deviation tell us much about a distribution of scores. In particular, they indicate where the center of the distribution is and how closely the scores cluster around the center. It's a fact worth remembering that a span of one standard deviation on either side of the mean covers approximately 68% of the scores in a **normal distribution.**

Earlier we determined that the hypothetical data we obtained with our Liberal–Conservative Scale revealed that students in the class we surveyed were, on the average, more conservative than students who took the national survey. The standard deviation shows, however, that there is considerable variation in opinion. In fact, several students (eight, to be exact) in our sample were *more* liberal than the national average. The resulting study is a correlational study.

Correlation: A Relationship Between Two Variables

Now let's take our research a step further by asking whether a person's tendency toward liberalism or conservatism is related to other personal characteristics. Do conservatives come from more affluent families? Are liberals more introverted? Do conservatives get better grades? Are liberals more likely to major in the social sciences and humanities, while conservatives major in business or the natural sciences? Such questions deal with **correlation,** which is a relationship between *variables*. The resulting study is a correlational study.

To illustrate, suppose we have a hypothesis stating that the conservative students at your school are more money-oriented than the liberal students. (This hypothesis may be true—or it may *not* be true. Only a scientific test can tell.) We can put our hypothesis to a test by first defining "money-orientation" as "expected earnings five years after graduation." Next, we would obtain a sample of students from your school and request two items of information from each of them: (a) how much money they expect to be making five years after graduation and (b) their score on our Liberal–Conservative Scale. Our hypothesis, then, would predict that scores on the LCS would be associated—or *correlated*—with expected income. Specifically, we would predict that lower income estimates would be associated with higher LCS scores, while higher expected incomes would come from respondents with lower LCS scores. An analysis of the data should reveal whether or not the hypothesis is true.

Normal distribution A bell-shaped curve, describing the spread of a characteristic throughout a population.

Correlation A relationship between variables, in which changes in one variable are reflected in changes in the other variable—as in the correlation between a child's age and height.

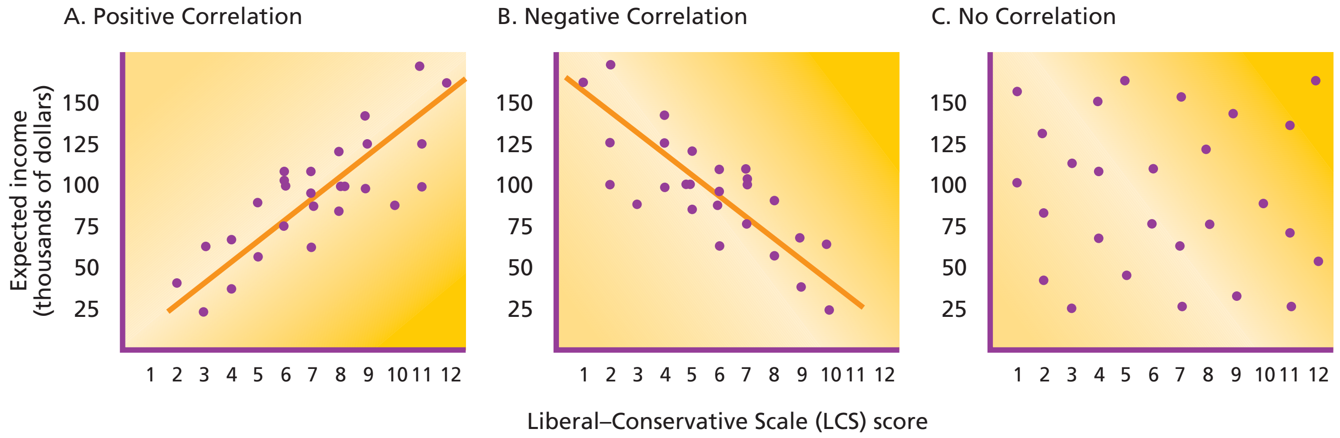


FIGURE 2.5 Three Types of Correlation

A. Positive Correlation; B. Negative Correlation; C. No Correlation

For certain, we will get one of three possible outcomes: a *positive correlation*, a *negative correlation*, or *no correlation* between expected income and LCS scores. Each of these possibilities is shown graphically for a class of 26 students in Figure 2.5. Note that if we find a positive correlation, then increasing LCS scores will be associated with increasing expected incomes. In this case, the points on the graph would cluster near an upward-sloping line, as in Figure 2.5A. If, however, the two variables turn out to have a negative correlation, then increasing LCS scores will be associated with decreasing income expectations—and the points on the graph will cluster near a downward-sloping line, as in Figure 2.5B. (A negative-correlation pattern was predicted by our hypothesis.) If there is no correlation (or a near-zero correlation), the dots will fall randomly all over the graph, as in Figure 2.5C.

We can tell most precisely which one of these relationships we have by looking at the **correlation coefficient,** a number that falls between –1.0 and +1.0. (You will also learn how to calculate this number in your introductory statistics class.) If people with high scores on one variable tend to have high scores on the other variable, the correlation is positive, and the correlation coefficient will also be positive (greater than 0). If, however, people with high scores on one variable tend to have low scores on the other variable, the correlation is negative, and the coefficient will also be negative (less than 0). If there is no consistent relationship between the scores, the correlation will be close to 0.

Making Inferences with Inferential Statistics

Now that we have seen how to use descriptive statistics to describe a set of data, let's turn to **inferential statistics,** which are used to determine (infer) whether the scores from two or more groups are essentially the same or different. For example, if you hypothesized that time spent studying is associated with the grades students receive, you could use inferential statistics to compare the average amount of study time in a sample of students with high grades to that of a sample with low grades. The details of the statistical tests we could use in this example are beyond the scope of this brief introduction to statistics. Suffice it to say that most inferential statistics take into account the differences between mean scores of each group, along with their standard deviations.

■ **Correlation coefficient** A number between –1 and +1 expressing the degree of relationship between two variables.

■ **Inferential statistics** Statistical techniques (based on probability theory) used to assess whether the results of a study are reliable or whether they might be simply the result of chance. Inferential statistics are often used to determine whether two or more groups are essentially the same or different.

To be sure that any differences you find are real, you must also factor in the size of the sample you used. As you might expect, with a small sample, a relatively large difference in grades between the two sample groups is required before you can conclude that the samples represent truly different populations. In addition, you must consider the distribution of scores in both groups. Do the sample scores approximate a normal distribution? If not, you may have to use alternative statistical tests—which you will learn about in a course on statistics.

Sampling To have confidence in your results you must, of course, make sure that your sample was selected in an unbiased manner. The safest way is to select participants at random, by a method such as drawing names from a hat. Sometimes obtaining a **random sample** is not practicable: Just imagine trying to get a random sample of all college students in the country! A good alternative is to take a **representative sample.** (This is what the Gallup Poll and other major polling services do.) A representative sample reflects the important variables in the larger population in which you are interested—variables such as age, income level, ethnicity, and geographic distribution. Remarkably, a carefully selected representative sample of only a few hundred persons is often sufficient for public-opinion pollsters to obtain a highly accurate reflection of the political opinions of the entire population of a country.

Statistical Significance A researcher who finds a difference between the mean scores for two sample groups must ask whether it occurred simply because of chance or whether it represents a real difference between the two populations from which the samples were drawn. To illustrate, suppose that we compare the mean scores on the Liberal–Conservative Scale for the men and women we surveyed. If gender has no influence on liberal–conservative attitudes, then we would expect the two means to be fairly similar, and any slight difference between the samples would be due to chance. This would most likely be the case, say, if we sampled 25 men and 25 women and their mean LCS scores differed by only .1 point.

But what if the difference between the scores for the two groups is somewhat larger—say, 3 points? As you learned earlier, less than a third of the scores in a normal distribution should be greater than one standard deviation above or below the mean. So, if there is no real difference between the men's group and the women's group, the chances of getting a male sample with a mean score that is more than, say, two standard deviations above or below the mean for the female sample would be very small. Thus, a researcher who does get a difference that great would feel fairly confident that the difference is a real one and is somehow related to gender. The actual computation required to demonstrate this takes the sample size, the size of the difference, and the spread of the scores into consideration. Again, the details of this computation are beyond the scope of this book, but they are not difficult.

By common agreement, psychologists accept a difference between the groups as "real" or "significant" when the probability that it might be due to chance is less than 5 in 100 (indicated by the notation $p < .05$). A **significant difference,** then, is one that meets this criterion. However, in some cases, even stricter probability levels are used, such as $p < .01$ (less than 1 in 100) and $p < .001$ (less than 1 in 1000).

As you can see, any conclusion drawn from inferential statistics is only a statement of the probability that the results reflect a real difference in the world, rather than a chance difference in the samples selected. Science is never about absolute certainty. Truth in science is always open to revision by later data from better studies, developed from better hypotheses and better samples.

■ **Random sample** A sample group of subjects selected by chance (without biased selection techniques).

■ **Representative sample** A sample obtained in such a way that it reflects the distribution of important variables in the larger population in which the researchers are interested–variables such as age, income level, ethnicity, and geographic distribution.

■ **Significant difference** Psychologists accept a difference between the groups as "real," or significant, when the probability that it might be due to an atypical sample drawn by chance is less than 5 in 100 (indicated by the notation $p < .05$)

PSYCHOLOGY IN YOUR LIFE: STATISTICS IN POLITICS

It is statistics that tell us what numbers mean. Each consumer product survey or opinion poll ends with a statement on how significant the data are and what they mean. In politics, surveys and statistics dictate how campaigns spend their money, where candidates spend their time, and (to some extent) what candidates say. By conducting surveys, candidates can find out the best way to phrase a message and learn to whom they should say it. By surveying a representative sample of her or his (potential) constituents, a candidate can test ideas and advertisements, and then modify them based on the results, before "going public" with platforms or ads.

The numbers that candidates get from the surveys are the key components here. Survey questions that demonstrate little or no significance (statistically) let the candidate know that the issue a question addresses is not important to constituents, and in a campaign this information is very important. A candidate talking about issues that are not relevant or important to their constituents is not a candidate who is going to win election! Candidates (and campaigns) often employ psychologists to write surveys and analyze their results to help tailor the message so that the candidates can spend their time and money talking about the issues that they—as well as the people they would like to represent—care about.

CHECK YOUR UNDERSTANDING

1. **ANALYSIS:** Which of the following correlation coefficients would a statistician know, at first glance, is a mistake?
   - **a.** 0.0
   - **b.** +1.1
   - **c.** +1.0
   - **d.** −0.7
   - **e.** −0.2

2. **RECALL:** Which of the following is a measure of central tendency?
   - **a.** mean
   - **b.** correlation
   - **c.** random sample
   - **d.** frequency distribution
   - **e.** histogram

3. **RECALL:** The simplest measure of variability is
   - **a.** mean.
   - **b.** median.
   - **c.** mode.
   - **d.** standard deviation.
   - **e.** range.

4. **ANALYSIS:** Most psychologists accept a difference between groups as "real," or significant, under which of the following conditions?
   - **a.** $p < .5$
   - **b.** $p < .3$
   - **c.** $p < .1$
   - **d.** $p < .05$
   - **e.** $p = 0$

ANSWERS: 1. b 2. a 3. e 4. d

USING PSYCHOLOGY TO LEARN PSYCHOLOGY

Research in Practice

Yumi and Maria learned all about the different research methods we have considered in this chapter. As we look at their study, we can see all of the essential aspects of research. They began with an abstract that summarizes the research and its findings. Following is their actual introduction. They conducted a lengthy discussion of the available research on the topic, culling their information from a variety of authors and current and relevant sources.

In the course of their literature review, Yumi and Maria discovered the best method to use (the survey) and developed their instrument, which was translated into Japanese for student use overseas. The report also describes their data collection and analysis methodologies; clearly identifies the items in their survey that they would change; and indicates how they would change them. Finally, the conclusion explains their reactions to what they discovered: "Surprisingly, both genders of the Japanese are less happy with their bodies although their BMI's are generally lower than [those of] Americans" (Kallman & Rydlun, 2002).

This study below shows all the hallmarks of good research. It is through conscientious investigation such as this that we have learned virtually everything we know about the natural world.

Japan vs. America: Differences in Health and Body Image

Maria Kallman and Yumi Rydlun

Walt Whitman High School

ABSTRACT

It has been shown that Japanese and American teenagers' lifestyles differ significantly in their level of physical activity and eating habits. In comparing Japanese and American teenagers' health habits and body image, conclusions can be drawn regarding the implications of lifestyle habits on body image. A multi-faceted survey was developed to assess lifestyle health habits and body image in a culturally non-biased way. Two hundred surveys were distributed to Japanese and American teenagers in their native languages. It was hypothesized that due to the different cultural factors in Japan such as a more balanced diet and more physical exercise, Japanese teenagers would be happier with their bodies. Data found that overall, the subjects' lifestyles did not differ significantly in health between the two countries, and, to further nullify the hypothesis, showed that the Japanese teenagers were less happy with their bodies.

INTRODUCTION/LIT REVIEW

Body image, or the way we perceive our physical selves, is generally at an all-time low during the teenage years (Sneddon, 1999). The transitional teenage years make teens especially susceptible to outside influences on their self-perception. In the United States, recent studies have shown that at least 50% of adolescent girls read fashion magazines such as *Seventeen* or *Vogue* regularly, and therefore are exposed to the media's projection of what is ideal (Sneddon, 1999). This ideal is becoming harder and harder to achieve. For example, in 1920 the popular film star and owner of the 'perfect body,' Anette Kellerman, was 5 feet 3 inches tall and weighed 130 pounds, with a healthy body mass index of 23. Today, renowned supermodel Kate Moss is 5 feet 7 inches and weighs 100 pounds, with a BMI of about 16 (Sneddon, 1999). In addition, a generation ago, the average model weighed 8% less than the average woman, whereas now, she weighs 23% less (Sneddon, 1999).

The unachievable nature of these body types causes teens to revert to dieting, their self-esteem drops, and they become prone to eating disorders. In addition to promoting an impossibly, often unhealthy body weight, magazines, television, and the numerous diet products available "can give girls the message that dieting is a necessary part of growing up as a woman" (Sneddon, 1999). Similarly, television and magazines favor muscular men with washboard stomachs. As boys hit puberty, they become aware of their stature. "In contrast to the girls, boys want to put on more weight than take it off" (Sneddon, 1999). Males can easily be embarrassed in the locker rooms, and may turn to steroid use to increase muscle mass. Dr. Gary Wadler, winner of a 1993 International Olympic Committee prize for his work on drugs and sports, related steroid takers to victims of eating disorders like anorexia and bulimia. "One is the relentless pursuit of bigness, the other is the relentless pursuit of thinness." (Sneddon, 1999). However, it has been found that western societies' excessive valuation of thinness does not generalize to other cultures, although focus on women's beauty is just as apparent. Another study suggests that white British females had more concern about appearance than black or Asian groups (Ramachandran, 1994).

One of the major differences between Japan and America's health is diet-related. Until recently, the Japanese were mostly free of heart disease, breast, and lung cancer, and still live longer than Americans (Brody, 1999). Fourteen or fifteen percent of Japanese daily calories are from fat, whereas groups such as the National Academy of Sciences and the American Heart Association have recommended that Americans reduce their consumption of fat so that it makes up a maximum of 30% of their daily caloric intake (Brody, 1997). In addition, a traditional Japanese diet, which includes many soybeans and vegetables, has been proven to raise levels of a compound called genistein which blocks angiogenesis, the growth of new blood vessels. Among other benefits, a diet with many vegetables like that of the Japanese lessens the risk of developing large tumors because tumors require angiogenesis in order to be fed (Brody, 1997). Furthermore, in addition, American's diets have been found to be much higher in red meat, while Japanese prefer fish. Red meat is higher in saturated fat than fish, and has many long-term negative effects on health, including detrimental effects on blood pressure and cholesterol. Yet although Japanese traditionally have eaten a healthier diet with more vegetables and fish, and less red meat and fast food, in the recent years the country has begun to adopt American fast food and eating habits.

A healthy lifestyle requires not only a smart diet but physical activity as well. Despite the national obsession with dieting, the *Los Angeles Times* writes that 50% of thirteen- to nineteen-year-olds in the U.S. do not engage in strenuous physical activity (Sneddon, 1999). A major difference in the activity levels of American teens and those throughout much of the rest of the world lies in their mode of transportation. Most American cities and suburbs are relatively spread out and roads are designed mainly for car travel, as opposed to Europe and Asia, where suburbs are less common and most parts of cities are walking or bike-riding distance from a person's residence. In addition, American teens can easily obtain a driver's license around age 16, whereas in Japan, a large fee is required and the minimum age is 18. Japanese city streets are also much more crowded with pedestrians and bikes and therefore it is difficult to navigate a car.

A sedentary lifestyle has been cited as responsible for as many as 250,000 deaths per year in the U.S. (Brody, 1999). Exercise has many benefits, such as boosting the immune system by increasing numbers of white blood cells, strengthening bones to ward off osteoporosis, increasing basal metabolic rate, and, perhaps most importantly, lowering the risk of heart disease. Low physical activity in high school students has been associated with negative behaviors, and adolescents who are overweight are at a greater risk of being overweight as adults (Risk, 2000). According to a recent study by the National Center for Health Statistics in the Centers for Disease Control and Prevention, the number of overweight American adults, which was stable until 1980, abruptly jumped to one-third of adults by 1991, and then to one-half by 1996 (Brody, 1997).

It is mostly in the psychological benefits of exercise, however, that lies the link between body image and health. Numerous studies have shown a positive correlation between a healthy lifestyle and positive self-worth. Correlational research has suggested that physical activity strongly improves the development of a positive self-concept, which encompasses many aspects of an individual's self-esteem including body image (Kazdin, 2000). Exercise also has a positive impact on a variety of psychological outcomes such as mood, symptoms of depression and anxiety, perceived stress, and psychological well-being, which are factors that have been linked to improvement in health-related quality of life (Kazdin, 2000).

Other factors of a healthy lifestyle include low amounts of stress, restricted use of alcohol and drugs, adequate amounts of sleep, and limited consumption of caffeine. Caffeine has a negative effect on calcium metabolism, which suppresses bone growth in adolescents (Brody, 1997).

Based on these findings, it can be inferred that a healthy lifestyle, including exercise and a well-balanced diet, is highly correlated with a positive body image. This study will investigate possible correlations between health and body image and how they may differ among Japanese and American teenagers. We hypothesize that Japanese teenagers eat a healthier diet, get more exercise, and lead a generally healthier lifestyle, and are therefore happier with their bodies.

METHODS

Participants

Surveys were randomly distributed to students in Japan and America. In Japan, two co-ed private high schools, one all-female private high school, and one university in the Tokyo and Odawara areas were surveyed. In the U.S., one co-ed private high-school and one co-ed public high school in the metropolitan Washington area were surveyed. All of the schools were composed of students of similar socioeconomic background. Approximately 50 surveys were distributed in Japanese to the Japanese schools, and about 70 English versions of the survey were distributed to the U.S. schools. Of the 129 U.S. surveys that were returned, 70 subjects were male and 59 were female. Of the 173 Japanese surveys returned, 57 were male and 116 were female.

Survey

We formulated a 26-question survey to address four main areas: eating habits, exercise and physical health, health and lifestyle habits, and body image (see survey on page 54).

Procedure

We converted the Japanese metric measurements of weight and height into inches and pounds, and then calculated each subject's BMI by using the following formula:

$$\frac{\text{weight in pounds}}{\text{height (inches squared)}} \times 703$$

We then calculated each subject's Ideal Variance Index, a measurement of the difference between subjects' actual ten-pound weight range and their ideal weight. For the questions answered with words, we assigned numerical values to each word choice. To calculate participation in sports, we assigned a 1 to each 'no' and a 2 for each 'yes' for both questions 15 and 17 (Do you play sports in school? Outside of school?). We deleted any subjects who left fields blank in analyzing each question.

Results

While our study collected data that examined a number of health and body image issues, significant data was found that pertain to the following areas. Overall, the BMIs of the Japanese were significantly lower than those of the Americans ($p < .05$), yet more U.S. teens are content with their weight than the Japanese ($p < .05$) (fig. 2). There is a negative correlation among all groups between BMI and ideal weight, which shows that heavier people want to lose more weight (correl. –0.602 for females; –0.44 for males) (fig. 3). Japanese females weigh themselves significantly more often than the U.S. females ($p < .05$) while the difference between males was insignificant. Overall, the Japanese are less satisfied with their bodies than the Americans ($p < .05$) (fig 11).

Other data suggest that Japanese eat more candy as snacks, whereas Americans eat more grains (fig. 4), yet American teenagers prefer water and soda while the Japanese drink more fruit juice and milk. Also, teenagers in the United States use more drugs and alcohol (fig. 5,6). The Japanese sleep an average of 42 minutes less than the U.S. teens ($p < .05$) (fig. 8). Altogether, males eat more red meat; Japanese males eat less than the U.S. males, but Japanese females eat more than the U.S. females (fig. 7). Japanese males are significantly more stressed than the U.S. males, while the Japanese females are significantly less stressed than the U.S. females ($p < .05$) (fig. 9). American teenagers participate in significantly more sports than Japanese teenagers ($p < .05$) (fig. 10).

Dieting practices were insignificant between Japan and America, and there was no significant difference of weighing frequency between males of Japan and the U.S.

ANALYSIS

Through our surveys, we were able to show that Japanese teenagers are smaller in general by using height and weight to calculate BMI. We were also able to find many significant differences between Japan and America in a variety of different areas as well as areas in which the two countries were similar. For example, despite previous data that suggested that Japan had a healthier diet, our data showed that our Japanese subjects' diets were similar to those of the U.S. teens surveyed.

Perhaps the most significant flaw in the experiment was making the weight ranges too large so that calculating BMI became somewhat inaccurate. Also, there was no range for a weight over 190 pounds, and no height range over seventy six, consequently there was no way of calculating the BMI of some subjects. Seventeen percent of U.S. males were labeled as 'undefined' and not included in the graphs and calculations. The final BMI for each subject is based on the mean of the subject's minimum possible BMI and the maximum possible BMI, calculated using the smallest weight and the largest height, and the largest weight and the smallest height of the range, respectively. This range of BMI was about 2 units. This also made the ideal variance index somewhat of a loose estimate of subjects' satisfaction with their weight.

Similarly, the range for question 6 (how many times per week do you eat fast food) was 1–3, while it should have been 1, 2, 3, etc. to be more accurate. To make a better comparison of stress and body image, a scale from 0 to 5 would have been more helpful on question 10 (are you stressed?) rather than a response of always, sometimes, or never.

In addition, in order to get a better understanding of physical activity levels in each country, the survey could have included a question pertaining to the subject's daily mode of transportation.

The study would have been more accurate with more surveys, especially because many subjects left answers blank, but limited time and resources prevented surveying more subjects. The Japanese left many more fields blank than the U.S. subjects. Fifteen percent of Japanese males and 34 percent of Japanese females were omitted from the BMI calculations because one or more necessary fields were left blank. In addition, 9.5% of Japanese females excluded question 20 (are you unhappy with your body?) and 8% of the Japanese females left question 21 blank (Do you diet?). The survey could have included a few more questions relevant to body image, and excluded some irrelevant ones.

DISCUSSION

Surprisingly, both genders of the Japanese high schoolers are less happy with their bodies although their BMIs are generally lower than the Americans. In addition, the Japanese females weigh themselves more often than their U.S. counterparts. Although it was hypothesized that Japanese would have a healthier diet and lifestyle, in fact, the two countries' diets were similar and seemed to have no correlation to body image. However, Americans indicate that they play more sports than the Japanese, which could be a reason why they have a better body image overall. However, the survey questions related to diet and exercise could have been written more clearly to get a more accurate perception of health. These results lead us to the conclusion that the cultures of Japan and America, as they pertain to health habits, do not differ very much in the teen years, but that the societal ideals in Japan might be thinner than in the U.S.

REFERENCES

Brody, J. (1999). *Book of Health.* New York: The New York Times Company.

Kazdin, A. (Ed). (2000). Exercise and Physical Activity. In *The encyclopedia of psychology* (Vol. 2, pp. 285–287). Oxford: Oxford University Press.

Kazdin, A (Ed). (2000). Sport Psychology. In *The encyclopedia of psychology* (Vol. 7, pp. 456–458). Oxford: Oxford University Press.

Ramachandran, V. (Ed). (1994). Body Image Disturbance. In *The encyclopedia of human behavior* (Vol. 1). San Diego: Academic Press.

Sneddon, P. (1999). *Body Image: A Reality Check.* New Jersey: Enslow Publishers.

U.S. Department of Health and Human Services. (2000). Risk Behaviors: Adolescent Health. Hyattsville, MD: MacKay, A., Fingerhut, L., and Duran, C.

SURVEY

Please circle one:

Are you: MALE FEMALE

How old are you? 13 14 15 16 17 18 19

How much do you weigh? (in pounds)

< 91 91–100 101–110 111–120 121–130 131–140

141–150 151–160 161–170 171–180 181–190 > 190

What is your height? (in inches)

< 60 61–63 64–66 67–69 70–72 73–75 > 76

Do you usually eat breakfast, lunch, and dinner? YES NO

Do you usually eat between 9 PM and 5 AM? YES NO

In which category is your most typical snack food?

___ Candy

___ Fruits/vegetables

___ Grains (bread, cereal, rice, crackers)

___ Protein (peanut butter, tofu, meat, fish)

___ Dairy (yogurt, cheese, milk)

Do you prefer: Soda Fruit Juice Water Milk

Do you drink caffeinated beverages regularly? (coffee, tea, soda)

YES NO

About how many times do you eat fast food per week?

0 1–3 4–6 7–10 10+

About how long does it take you to eat dinner? (in minutes)

< 10 11–20 21–30 31–40 41–50 51–60 61+

About how often do you eat red meat per week?

0 1 2 3 4 5 6 7+

Do you play on a school sports team? YES NO

If yes, how many hours per week do you practice?

6–8 9–11 12–14 15–17 18–20 20+

Do you participate in sports not affiliated with school? YES NO

If yes, what? ____

How many hours per week do you practice? ________________________

How many hours of sleep do you get on school nights?

3 4 5 6 7 8 9+

Do you feel stressed? always sometimes never

Do you smoke cigarettes daily? YES NO

If yes, how many packs? .5 1 1.5 2+

Do you consume alcohol? often sometimes never

Do you do other drugs recreationally? often sometimes never

Are you UNhappy with your body? often sometimes never

Do you diet to change your body? often sometimes never

What is your ideal weight? _______________

How often do you weigh yourself?

___ Everyday ___ Once every two weeks

___ More than once a week ___ More than once a month

___ Once a week ___ Hardly ever

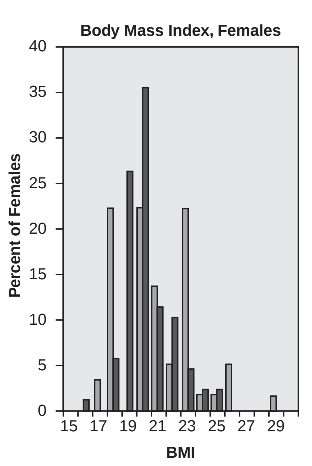


FIGURE 1

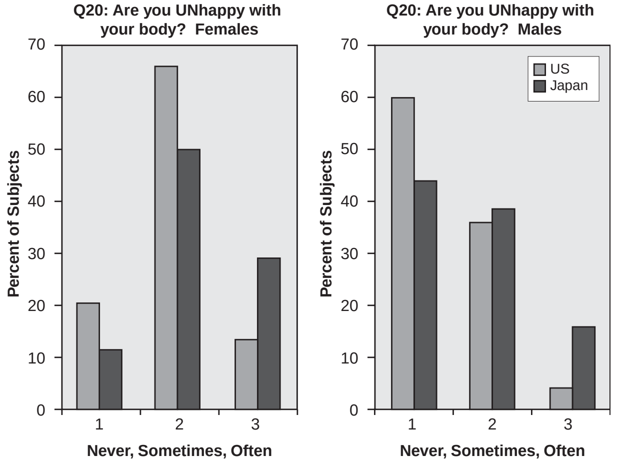


FIGURE 2

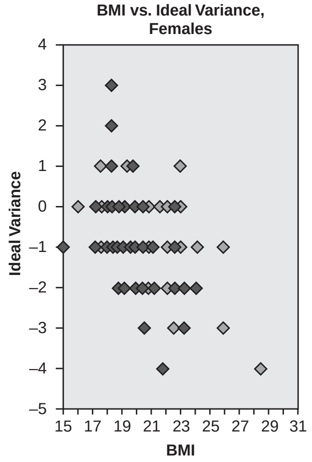


FIGURE 3

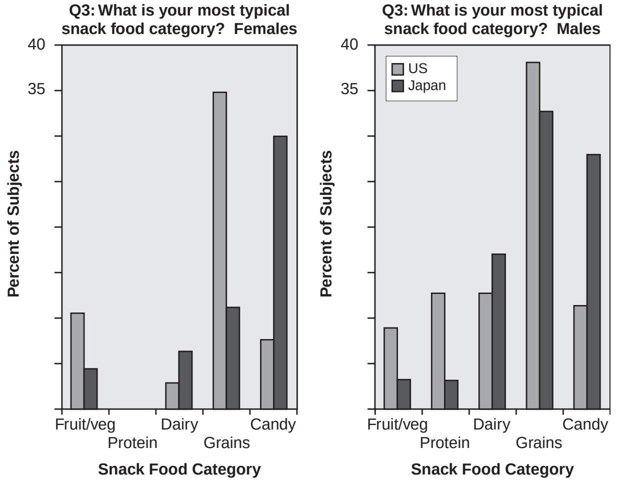


FIGURE 4

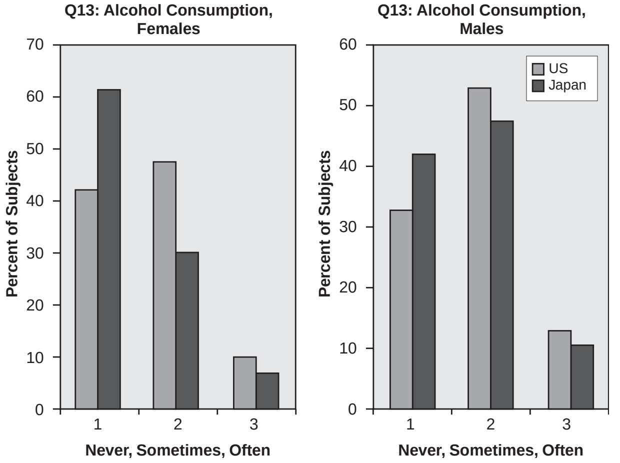
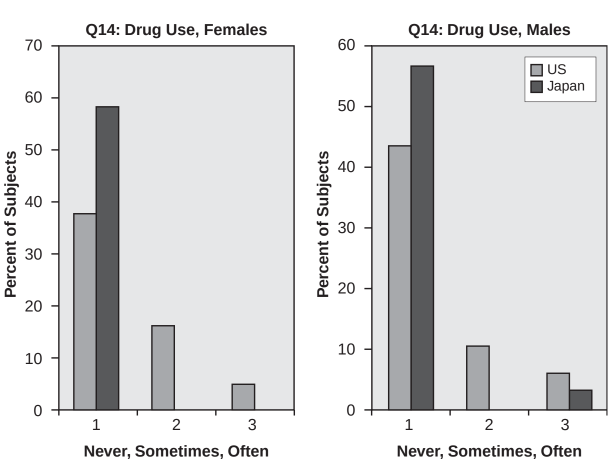
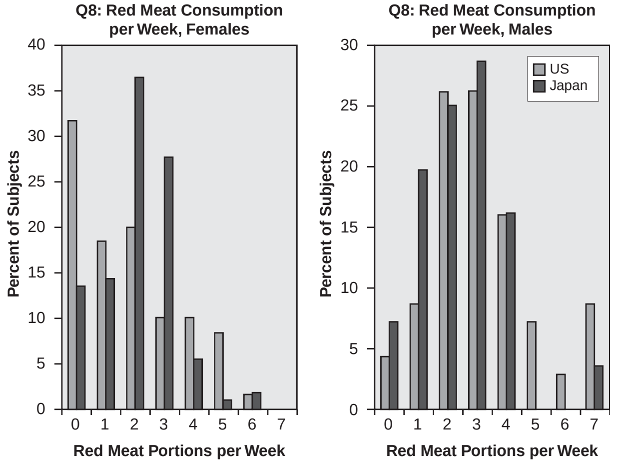


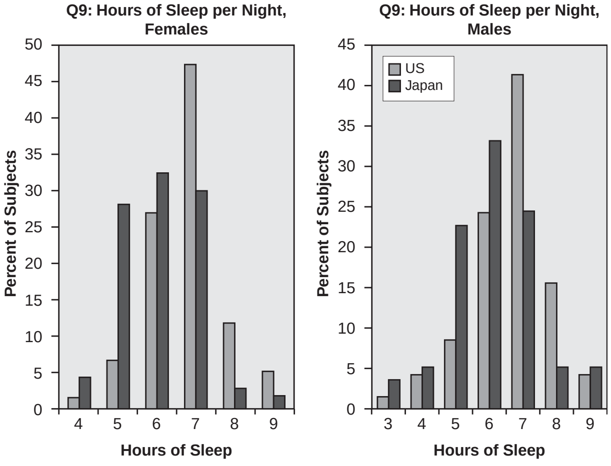
FIGURE 5



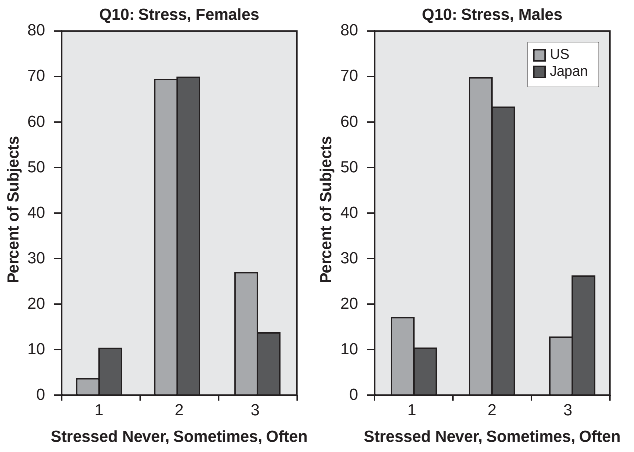
● **FIGURE 6**



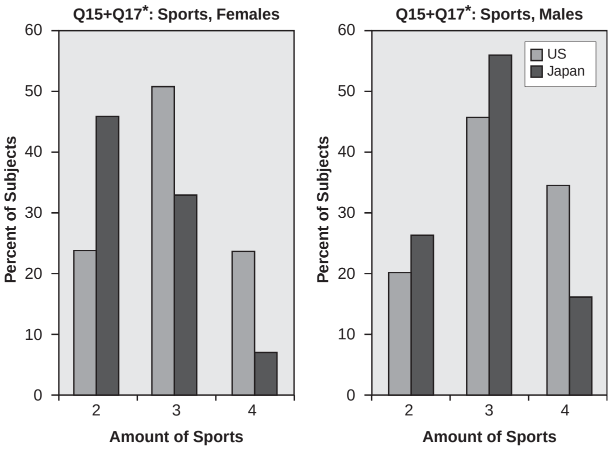
● **FIGURE 7**



● **FIGURE 8**

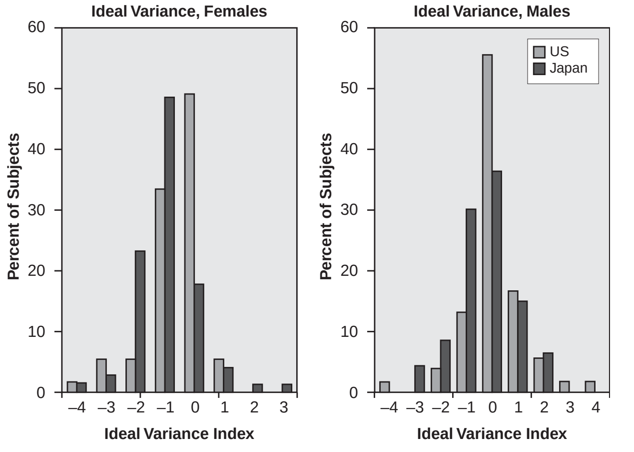


● **FIGURE 9**



*For Sports
Q15 Do you play sports in school?
Q17 Do you play sports outside of school?
} For each "yes" → 2 marks
For each "no" → 1 mark

● **FIGURE 10**



● **FIGURE 11**

CHAPTER SUMMARY

HOW DO PSYCHOLOGISTS DEVELOP NEW KNOWLEDGE?

Psychologists use many different methods to conduct research and gain new knowledge. In an experiment, the steps in performing an experiment are designed to produce results that are replicable. Although they may seem laborious, they are followed so that any result can be either proven or refuted. In contrast to the experiment, the seven quasi-experimental methods do not demonstrate cause and effect, but rather, demonstrate how certain phenomena are related.

All research is conducted in such a way to eliminate any sources of bias be they confounding or extraneous variables. Even though it was discussed last, ethical considerations are the cornerstone of all research. No experiment or study can be conducted without the approval of a research panel such as an IRB or an IACUC.

Psychologists, like researchers in all other sciences, use the scientific method to test their ideas empirically.

HOW DO WE MAKE SENSE OF THE DATA?

By organizing and analyzing data, psychologists (and all researchers) can draw conclusions about the data they collect. Descriptive statistics help us describe the main characteristics of the data such as mean, median, and mode, as well as finding the spread of the data and the relationship between variables, as in a correlation.

Inferential statistics let us know what the data mean, and allow us to make decisions based on the data and its significance. Data is considered to be significant when the probability that it might be due to chance is less that 5 in 100 ($p < .05$). Some studies may have stricter criteria.

Researchers use statistics for two major purposes: (1) descriptively to characterize measurements made on groups or individuals and (2) inferentially to judge whether those measurements are the result of chance.

REVIEW TEST

For each of the following items, choose the single best answer. The correct answers appear at the end.

1. Which of the following is an aspect of an experiment that the experimenter *cannot* control?
   a. controls
   b. dependent variable
   c. extraneous variables
   d. hypothesis
   e. independent variable

2. Which experimental method establishes cause and effect?
   a. cohort-sequential study
   b. correlational study
   c. experiment
   d. ex post facto design
   e. survey

3. Which of the following is *not* a step in the "ethics cascade"?
   a. *Who should decide* what is morally justifiable in the conduct of research?
   b. Are controlled research studies *ever* necessary or appropriate?
   c. Should all research have a *foreseeable practical benefit?*
   d. Who should conduct research?
   e. What *specific topics* are worthy of research?

4. The mean is the
   a. average.
   b. range.
   c. middle score.
   d. most frequently occurring score.
   e. standard deviation.

5. Which type of study is the "next best thing" to a longitudinal study?
   a. cross-sectional study
   b. cohort-sequential study
   c. ex post facto design
   d. naturalistic observation
   e. survey

6. What do researchers use to summarize, describe, and analyze the results of their research?
   a. case studies
   b. experiments
   c. naturalistic observations
   d. statistics
   e. surveys

7. The primary purpose of a survey is to
   a. describe an entire population.
   b. determine cause-and-effect relationships.
   c. discover attitudes and beliefs.
   d. find correlations between variables.
   e. test hypotheses.

8. A survey would be most useful in determining
   a. how common alcohol abuse is among high school students.
   b. how people develop over time.
   c. whether praise leads to more long-lasting changes in behavior than punishment.
   d. the effect of laughter on illness.
   e. whether rats run mazes faster in darkness or in light.

9. Sample is to population as
   a. child is to adult.
   b. large is to small.
   c. not representative is to representative.
   d. part is to whole.
   e. valid is to invalid.

10. Which of the following correlation coefficients indicates the strongest link between two variables?
   a. +0.4
   b. 0.0
   c. –0.4
   d. 0.8
   e. –0.9

ANSWERS: 1. c 2. c 3. d 4. a 5. b 6. d 7. c 8. a 9. d 10. e

KEY TERMS

Scientific method (p. 28)
Empirical investigation (p. 28)
Theory (p. 28)
Hypothesis (p. 29)
Operational definition (p. 30)
Independent variable (IV) (p. 30)
Random presentation (p. 30)
Data (p. 31)
Dependent variable (DV) (p. 31)
Replicate (p. 32)
Experiment (p. 33)
Confounding or **extraneous variables** (p. 33)
Controls (p. 33)
Random assignment (p. 34)
Ex post facto (p. 34)
Correlational study (p. 34)
Survey (p. 35)
Naturalistic observation (p. 35)
Longitudinal study (p. 35)
Cross-sectional study (p. 35)
Cohort-sequential study (p. 35)
Personal bias (p. 36)
Expectancy bias (p. 36)
Double-blind study (p. 36)
Institutional Review Board (IRB) (p. 37)
Institutional Animal Care and Use Committee (IACUC) (p. 37)
ABCs of laboratory animal research (p. 37)
Frequency distribution (p. 44)
Histogram (p. 44)
Descriptive statistics (p. 44)
Mean (p. 45)
Median (p. 46)
Mode (p. 46)
Range (p. 46)
Standard deviation (SD) (p. 46)
Correlation (p. 47)
Normal distribution (p. 47)
Correlation coefficient (p. 48)
Inferential statistics (p. 48)
Random sample (p. 49)
Representative sample (p. 49)
Significant difference (p. 49)

AP* REVIEW: VOCABULARY

Match each of the following vocabulary terms to its definition.

1. Cohort-sequential study
2. Dependent variable
3. Double-blind study
4. Empirical investigation
5. Hypothesis
6. Independent variable
7. Operational definitions
8. Random assignment
9. Replication
10. Standard deviation (SD)

_____ **a.** A stimulus condition so named because the experimenter changes it independently of all the other carefully controlled experimental conditions.

_____ **b.** This type of study takes a cross section of the population and then follows each cohort for a short period of time.

_____ **c.** An approach to research that relies on sensory experience and observation for research data.

_____ **d.** An explanation of how concepts are to be measured or what operations are being employed to produce them.

_____ **e.** A measure of variability that indicates the average difference between the scores and their mean.

_____ **f.** Doing a study over to see whether the same results are obtained.

_____ **g.** Ensures that each subject in an experiment must have an equal chance of being in the experimental group or the control group.

_____ **h.** A statement predicting the outcome of a scientific study; a statement describing the relationship among variables in a study.

_____ **i.** An experimental procedure in which neither the participants nor the experimenters working directly with them know what effects the independent variable may have or to which participants it is being administered.

_____ **j.** The measured outcome of a study; the responses of the subjects in a study.

AP* REVIEW: ESSAY

Use your knowledge of the chapter concepts to answer the following essay question.

Design an experiment to see how development changes over time. Your response should include two experimental methods and should enumerate the pros and cons of each. Be sure that your response includes

a. Two methods, with pros and cons of each
b. Subject selection
c. Methodology
d. Means for analyzing data
e. Drawing a conclusion
f. Discussion of any relevant ethical issues

OUR RECOMMENDED BOOKS AND VIDEOS

BOOKS

Abramson, Charles I. (1990). *Invertebrate learning: A source book*. Washington, D.C.: APA. This spiral-bound book provides a plethora of experiments and information on learning in invertebrates. It discusses methods as well as ethics and specific experiments.

Akins, Chana K., Panicker, S. and Cunningham, C. (eds.) (2004). *Laboratory animals in research and teaching: Ethics, care, and methods*. Washington, D.C.: APA. This book by a collection of authors discusses animals in teaching and research in universities, colleges and high schools. There is an excellent section on ethics.

Blatner, David (1999). *The joy of pi*. New York: Walker & Co. Kids love it because it is loaded with neat graphics, histories, and fun facts about the world's favorite number.

Jacobs, Harold R. (1994). *Mathematics—a human endeavor*. New York: W.H. Freeman & Co. A great book that connects math to the world around us.

Guide for the Care and Use of Laboratory Animals, National Research Council (1994), National Academy Press. This book describes the humane care and use of animals for institutions of all sizes and resources. An absolute must for anyone considering research.

VIDEOS

Stand and Deliver. A great film about teaching mathematics (calculus) in inner city Los Angeles, and the trials and tribulations of the students and an extraordinary teacher (Jaime Escalante). (*Rating PG*)

The Andromeda Strain. Film about a virus that lands in Arizona, and researchers from labs across the United States rush to fight and destroy it before millions die. (*Rating G*)

Mon Oncle d'Amérique. The behavioral theories of Henri Laborit are discussed in this comedy/drama set as a pseudo-documentary. Well acted. The film is in French with English subtitles. Winner of six French Caesar Awards. (*Rating PG*)