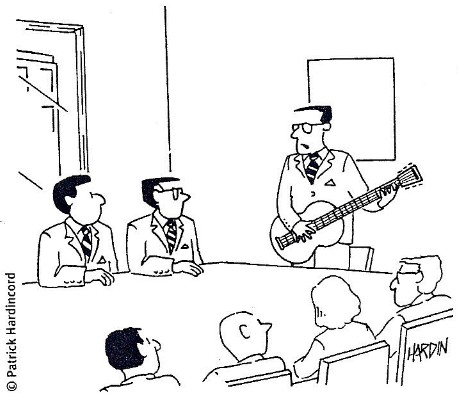

© Patrick Hardincord

"Figures can be misleading—so I've written a song which I think expresses the real story of the firm's performance this quarter."

Statistical Reasoning

Preview: Having gathered data, we must next organize, summarize, and make inferences from it, using statistics. Today's statistics are tools that help us see and interpret what the unaided eye might miss.

Off-the-top-of-the-head estimates often misread reality and then mislead the public. Someone throws out a big round number. Others echo it and before long the big round number becomes public misinformation. A few examples:

- *One percent of Americans (2.7 million) are homeless.* Or is it 300,000, an earlier estimate by the federal government? Or 600,000, the estimate by the Urban Institute (Crossen, 1994)?
- *Ten percent of people are lesbians or gay men.* Or is it 2 to 3 percent, as suggested by various national surveys (Chapter 12)?
- *We ordinarily use but 10 percent of our brain.* Or is it closer to 100 percent? (Which 90 percent, or even 10 percent, would you be willing to sacrifice?) (Chapter 2)

The point to remember: Doubt big, round, undocumented numbers. Rather than swallow top-of-the-head estimates, focus on thinking smarter by applying simple statistical principles to everyday reasoning.

Answer to question on page 40: In the subliminal tapes experiment, the primary independent variable was the type of subliminal message, self-esteem versus memory. (This experiment actually had a second independent variable as well: people's beliefs about which tape they received.) The primary dependent variable was improvement on the self-esteem and memory measures.

Describing Data

Once researchers have gathered their raw data, their first task is to *organize* it. One way is to use a simple *bar graph*, as in **FIGURE 1.11**, which displays a distribution of trucks of different brands still on the road after a decade. When reading statistical graphs such as this, take care. Depending on what people want to emphasize, they can design the graph to make a difference look small or big. So think smart: When viewing figures in magazines and on television, read the scale labels and note their range.

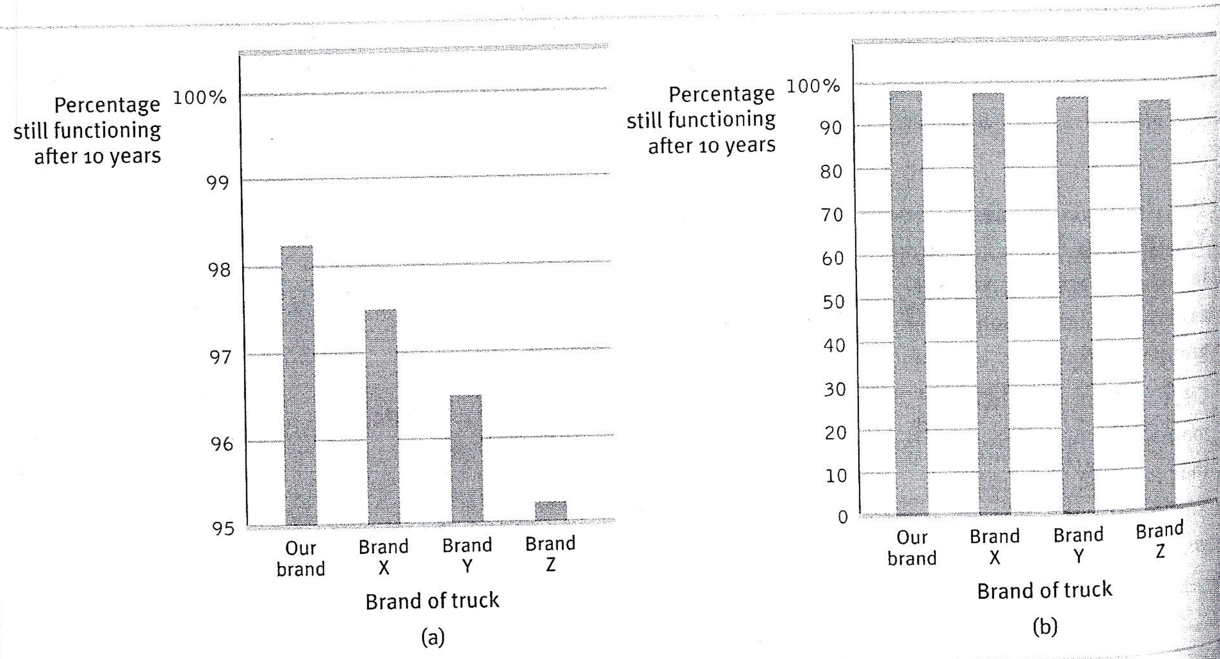

FIGURE 1.11
Read the scale labels
An American truck manufacturer offered a graph (a)—with actual brand names included—to suggest the much greater durability of its trucks. Note, however, how the apparent difference shrinks as the vertical scale changes (graph b).

Measures of Central Tendency

The next step is to summarize the data using the three measures of "central tendency." The simplest measure is called the **mode**, the most frequently occurring score. The most commonly reported is the **mean**, or arithmetic average—the total sum of all the scores divided by the number of scores. On a divided highway, the median is the middle. So, too, with data: The **median** is the middle score—the 50th percentile; if you arrange all the scores in order from the highest to the lowest, half will be above the median and half will be below it.

Measures of central tendency neatly summarize data. But consider what happens to the mean when a distribution is lopsided or *skewed*. With income data, for example, the mode, median, and mean often tell very different stories (**FIGURE 1.12**). This is because the mean is biased by a few extreme scores. When Microsoft founder Bill Gates sits down in an intimate cafe, its average (mean) patron instantly becomes a billionaire. Understanding this, you can see how a British newspaper could accurately run the headline "Income for 62% Is Below Average" (Waterhouse, 1993). Because the bottom *half* of British income earners receive only a *quarter* of the national income cake, most British people, like most people everywhere, make less than the mean. Professional athletes' incomes also form skewed distributions. In 1998, 66 percent of the National Basketball Association's 411 players made less than the average (mean) player salary (DuPree, 1998). The average ($2.24 million) was, of course, inflated by a few superstar salaries, led by Michael Jordan's $33.14 million.

The point to remember: Always note which measure of central tendency is reported. Then, if it is a mean, consider whether a few atypical scores could be distorting it.

mode the most frequently occurring score in a distribution.

mean the arithmetic average of a distribution, obtained by adding the scores and then dividing by the number of scores.

median the middle score in a distribution; half the scores are above it and half are below it.

range the difference between the highest and lowest scores in a distribution.

The average adult has one ovary and one testicle.

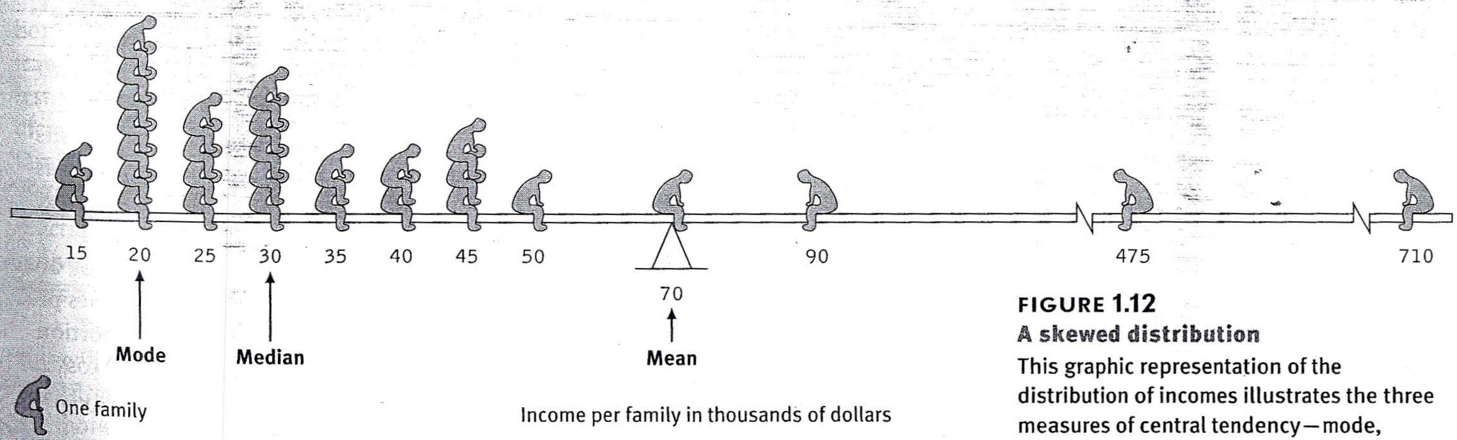

FIGURE 1.12
A skewed distribution
This graphic representation of the distribution of incomes illustrates the three measures of central tendency—mode, median, and mean. Note how just a few high incomes make the mean—the fulcrum point that balances the incomes above and below—deceptively high.

Measures of Variation

Knowing the value of an appropriate measure of central tendency can tell us a great deal. But it also helps to know something about the amount of *variation* in the data—how similar or diverse the scores are. Averages derived from scores with low variability are more reliable than averages based on scores with high variability. Consider a basketball player who scored between 13 and 17 points in each of her first 10 games in a season. Knowing this, we would be more confident that she would score near 15 points in her next game than if her scores had varied from 5 to 25 points.

The **range** of scores—the gap between the lowest and highest scores—provides only a crude estimate of variation because a couple of extreme scores in an otherwise uniform group, such as the $475,000 and $710,000 incomes in Figure 1.12, will create a deceptively large range.

standard deviation a computed measure of how much scores vary around the mean score.

statistical significance a statistical statement of how likely it is that an obtained result occurred by chance.

The more useful measure of how much scores deviate from one another is the **standard deviation.** It better gauges whether scores are packed together or dispersed, because it uses information from each score. (The computation assembles information about how much individual scores differ from the mean.) If your college or university attracts students of a certain ability level, their intelligence scores will have a smaller standard deviation than the one found in the more diverse community population outside your school.

Making Inferences

Data are "noisy." One group's average score (breast-fed babies' intelligence scores) could conceivably differ from another's (the formula-fed babies') not because of any real difference but merely due to chance fluctuation in the people sampled. How confidently, then, can we infer that an observed difference accurately estimates the true difference?

When Is an Observed Difference Reliable?

In deciding when it is safe to generalize from a sample, we should keep three principles in mind. Let's look at each in turn.

1. ***Representative samples are better than biased samples.*** As we have noted, the best basis for generalizing is not from the exceptional and memorable cases one finds at the extremes but from a representative sample of cases. No research involves a representative sample of the whole human population. Thus, it pays to keep in mind what population a study has sampled.
2. ***Less-variable observations are more reliable than those that are more variable.*** As we noted in the example of the basketball player whose points scored were consistent, an average is more reliable when it comes from scores with low variability.
3. ***More cases are better than fewer.*** An eager prospective university student visits two college campuses, each for a day. At the first, the student randomly attends two classes and discovers both instructors to be witty and engaging. At the next campus, the two sampled instructors seem dull and uninspiring. Returning home, the student tells friends about the "great teachers" at the first school, and the "bores" at the second. Again, we know it but we ignore it: Small samples provide less reliable estimates of the average than do large samples. The proportion of heads in samples of 10 coin tosses varies more than in samples of 100 tosses. Said differently, *averages based on many cases are more reliable* (less variable) than averages based on only a few cases.

The point to remember: Don't be overly impressed by a few anecdotes. Generalizations based on a few unrepresentative cases are unreliable.

"The poor are getting poorer, but with the rich getting richer it all averages out in the long run."

When Is a Difference Significant?

We can justifiably have the most confidence when we generalize from samples that (1) are representative of the population we wish to study, (2) give us consistent rather than highly variable data, and (3) are large rather than small. These principles extend to the inferences we make about differences between groups—as when we generalize from a gender difference in grades in our sample to the whole campus population.

Statistical tests help us determine significance by indicating the reliability of differences. Here is the logic behind them: When *averages* from two samples are each *reliable* measures of their respective populations (as when each is based on many ob-

PEANUTS

servations that have small variability), then their difference (sometimes even a very small difference) is likely to be reliable as well. (The less the variability in women's and men's aggression scores, the more confidence we would have that any observed difference is reliable.) But when the *difference* between the sample averages is *large*, we have even more confidence that the difference between them reflects a real difference in their populations.

In short, *when the sample averages are reliable and the difference between them is relatively large,* we say the difference has **statistical significance**. This simply means that the difference we observed is probably not due to chance variation between the samples. In judging statistical significance, psychologists are conservative. They are like juries who must presume innocence until guilt is proven. For most psychologists, proof beyond a reasonable doubt means not making much of a finding unless the odds of its occurring by chance are less than 5 percent (an arbitrary criterion).

When reading about research, you should remember that, given large enough or homogeneous enough samples, a difference between them may be "statistically significant" yet have little practical significance. For example, comparisons of intelligence test scores among several hundred thousand first-born and later-born individuals indicate that there is a highly significant tendency for first-born individuals within a family to have higher average scores than their later-born siblings (Zajonc & Markus, 1975). But because the scores differ by only one or two points, the difference has little practical importance. Such findings have caused some psychologists to advocate alternatives to significance testing (Hunter, 1997). Better, they say, to use other ways to express a finding's magnitude and reliability.

The point to remember: Statistical significance indicates the *likelihood* that a result will happen by chance. It does not indicate the *importance* of the result.

Using the principles discussed in this chapter will help us to think critically—to see more clearly what we might otherwise miss or misinterpret and to generalize more accurately from our observations. We do think smarter when we understand and use the principles of research methods and statistics (Fong & others, 1986; Lehman & others, 1988; VanderStoep & Shaughnessy, 1997). It requires training and practice, but developing clear and critical thinking abilities is part of your becoming an educated person. The report of the Project on Redefining the Meaning and Purpose of Baccalaureate Degrees (1985) eloquently asserts why there are few higher priorities in a college education:

> If anything is paid attention to in our colleges and universities, thinking must be it. Unfortunately, thinking can be lazy. It can be sloppy. . . . It can be fooled, misled, bullied. . . . Students possess great untrained and untapped capacities for logical thinking, critical analysis, and inquiry, but these are capacities that are not spontaneous: They grow out of wide instruction, experience, encouragement, correction, and constant use.

REVIEW AND REFLECT

Statistical Reasoning

To be an educated person today is to be able to apply simple statistical principles to everyday reasoning. One needn't remember complicated formulas to think more clearly and critically about data.

From this section's consideration of how we can organize, summarize, and make inferences from data—by constructing distributions and computing measures of central tendency, variation, and statistical significance—we derived five points to remember:

1. Doubt big, round, undocumented numbers.
2. When looking at statistical graphs in books and magazines and on television ads and news broadcasts, think critically: Always read the scale labels and note their range.
3. Always note which measure of central tendency is reported. Then, if it is a mean, consider whether a few atypical scores could be distorting it.
4. Don't be overly impressed by a few anecdotes. Generalizations based on only a few cases are unreliable.
5. Statistical significance indicates the *likelihood* that a result will occur by chance. It does not indicate the importance of the result.

CHECK YOURSELF: Consider a question posed by Christopher Jepson, David Krantz, and Richard Nisbett (1983) to University of Michigan introductory psychology students:

> *The registrar's office at the University of Michigan has found that usually about 100 students in Arts and Sciences have perfect marks at the end of their first term at the University. However, only about 10 to 15 students graduate with perfect marks. What do you think is the most likely explanation for the fact that there are more perfect marks after one term than at graduation?*

ASK YOURSELF: Find a graph in a popular magazine ad. How has the advertiser used (or abused) statistics to make a point?

Answers to the Check Yourself questions can be found in the yellow appendix at the end of the book.

Frequently Asked Questions About Psychology

Preview: A scientific approach can restrain our flawed intuition while satisfying our curiosity about what predicts or causes behavior. But for many, the idea of applying science to human affairs raises concerns about how well experiments relate to life, whether they apply to all cultures and both genders, how experimenters treat human and animal subjects, and how psychologists' values influence their work and its applications.

We have seen how case studies, surveys, and naturalistic observations help us describe behavior. We have also noted that correlational studies assess the relationship between two factors, which indicates how well, knowing one thing, we can