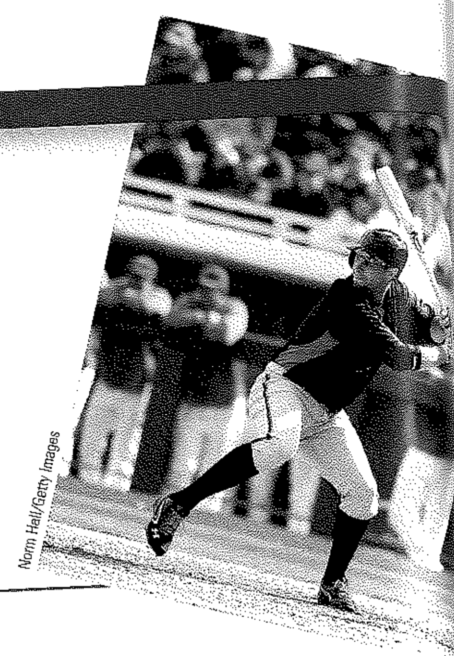
Module 7


Norm Hall/Getty Images

Statistical Reasoning in Everyday Life

Module Learning Objectives

7-1 Describe the three measures of central tendency, and discuss the relative usefulness of the two measures of variation.

7-2 Explain how we know whether an observed difference can be generalized to other populations.

In descriptive, correlational, and experimental research, statistics are tools that help us see and interpret what the unaided eye might miss. Sometimes the unaided eye misses badly. Researchers invited 5522 Americans to estimate the percentage of wealth possessed by the richest 20 percent in their country (Norton & Ariely, 2011). Their average person's guess—58 percent—"dramatically underestimated" the actual wealth inequality. (The wealthiest 20 percent possess 84 percent of the wealth.)

FYI

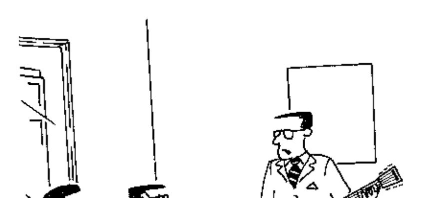
Asked about the *ideal* wealth distribution in America, Democrats and Republicans were surprisingly similar. In the Democrats' ideal world, the richest 20 percent would possess 30 percent of the wealth. The Republicans' ideal world was similar, with the richest 20 percent possessing 35 percent of the wealth. (Norton & Ariely, 2011).

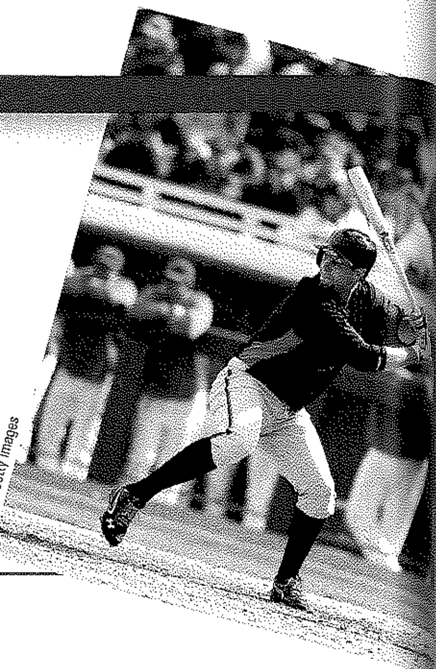
The Need for Statistics

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

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:

- *Ten percent of people are lesbians or gay men.* Or is it 2 to 3 percent, as suggested by various national surveys (Module 53)?
- *We ordinarily use but 10 percent of our brain.* Or is it closer to 100 percent (Module 12)?





statistics are tools that help us see
mes the unaided eye misses badly.
entage of wealth possessed by the
. Their average person's guess—58
lth inequality. (The wealthiest 20

pe an educated person today is to
easoning. One needn't memorize
about data.
ality and then mislead the public.
it, and before long the big, round

that women taking a particular contraceptive pill had a 100 percent increased risk of blood clots that could produce strokes. This caused thousands of women to stop taking the pill, leading to a wave of unwanted pregnancies and an estimated 13,000 additional abortions (which also are associated with increased blood clot risk). And what did the study find? A 100 percent increased risk, indeed—but only from 1 in 7000 women to 2 in 7000 women. Such false alarms underscore the need to teach statistical reasoning and to present statistical information more transparently.

Descriptive Statistics

7-1 How do we describe data using three measures of central tendency, and what is the relative usefulness of the two measures of variation?

Once researchers have gathered their data, they may use **descriptive statistics** to organize that data meaningfully. One way to do this is to convert the data into a simple *bar graph*, called a **histogram,** as in **FIGURE 7.1,** which displays a distribution of different brands of trucks still on the road after a decade. When reading statistical graphs such as this, take care. It's easy to design a graph to make a difference look big (Figure 7.1a) or small (Figure 7.1b). The secret lies in how you label the vertical scale (the *y-axis*).

The point to remember: Think smart. When viewing figures in magazines and on television, read the scale labels and note their range.

Measures of Central Tendency

The next step is to summarize the data using some *measure of central tendency,* a single score that represents a whole set of scores. The simplest measure is the **mode,** the most frequently occurring score or scores. 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 midpoint—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. In a symmetrical, bell-shaped distribution of scores, the mode, mean, and median scores may be the same or very similar.

descriptive statistics numerical data used to measure and describe characteristics of groups. Includes measures of central tendency and measures of variation.

histogram a bar graph depicting a frequency distribution.

mode the most frequently occurring score(s) 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.

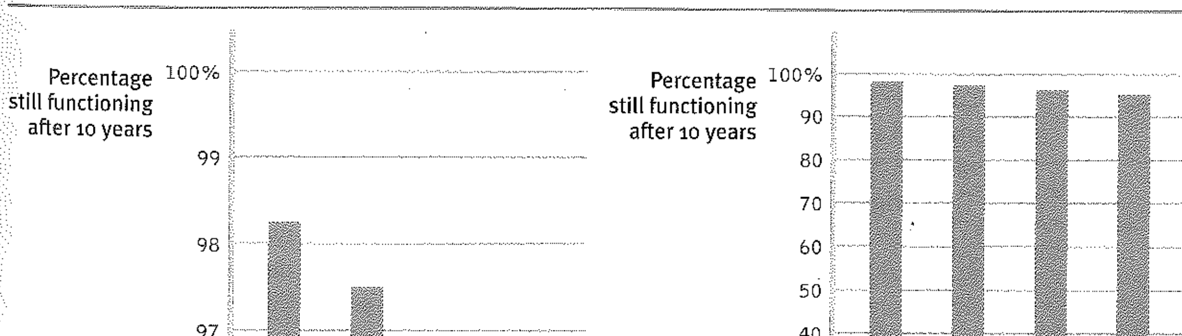


Figure 7.1
Read the scale labels An American truck manufacturer offered graph (a)—with actual brand names included—to suggest the much greater durability of its trucks. Note, however, how the apparent

Measures of central tendency neatly summarize data. But consider what happens to the mean when a distribution is lopsided, or **skewed,** by a few way-out scores. With income data, for example, the mode, median, and mean often tell very different stories (**FIGURE 7.2**). This happens because the mean is biased by a few extreme scores. When Microsoft co-founder Bill Gates sits down in an intimate café, its average (mean) customer instantly becomes a billionaire. But the customers' median wealth remains unchanged. 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. Mean and median tell different true stories.

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

FYI

The average person has one ovary and one testicle.

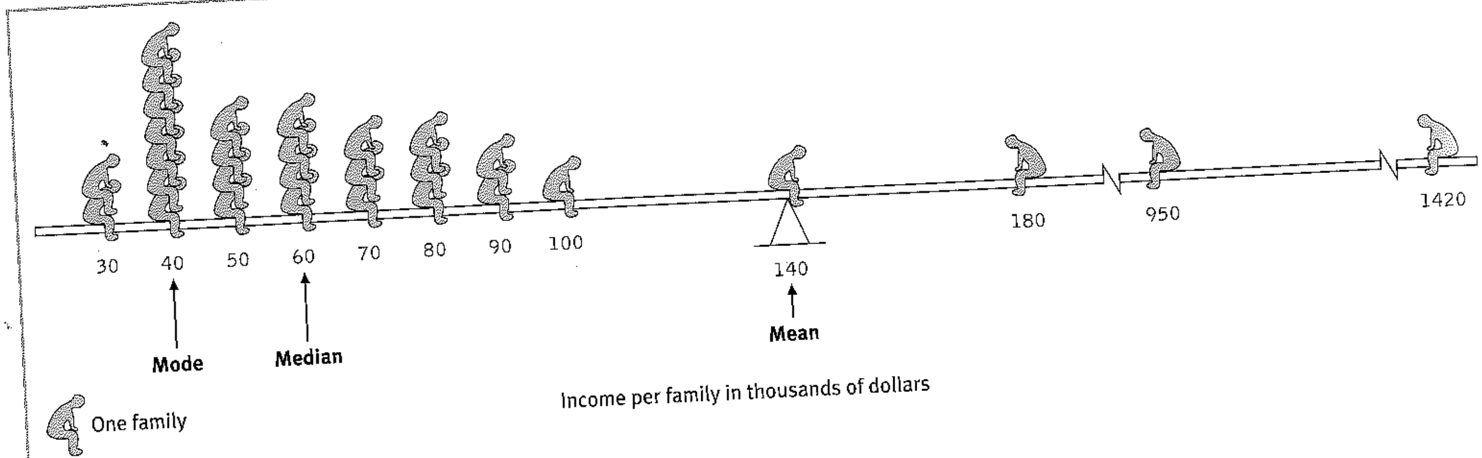


Figure 7.2

A skewed distribution This graphic representation of the distribution of a village's 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 the single number omits other information. It 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. A couple of extreme scores in an otherwise uniform group, such as the $950,000 and $1,420,000 incomes in Figure 7.2, will create a deceptively large range.

...ful standard for measuring how much scores deviate from one another ta. But consider what happens to , by a few way-out scores. With n often tell very different stories d by a few extreme scores. When café, its average (mean) customer dian wealth remains unchanged. er could accurately run the head- 1993). Because the bottom *half* of tional income cake, most British he mean. Mean and median tell

f central tendency is reported. If it e distorting it.

endency can tell us a great deal. But now something about the amount are. Averages derived from scores sed on scores with high variabil- and 17 points in each of her first onfident that she would score near rom 5 to 25 points.

nd highest scores—provides only a an otherwise uniform group, such create a deceptively large range.

scores deviate from one another

Table 7.1 Standard Deviation Is Much More Informative Than Mean Alone

Note that the test scores in Class A and Class B have the same mean (80), but very different standard deviations, which tell us more about how the students in each class are really faring.

Test Scores in Class A			Test Scores in Class B		
Score	**Deviation from the Mean**	**Squared Deviation**	**Score**	**Deviation from the Mean**	**Squared Deviation**
72	−8	64	60	−20	400
74	−6	36	60	−20	400
77	−3	9	70	−10	100
79	−1	1	70	−10	100
82	+2	4	90	+10	100
84	+4	16	90	+10	100
85	+5	25	100	+20	400
87	+7	49	100	+20	400
Total = 640		Sum of (deviations)² = 204	Total = 640		Sum of (deviations)² = 2000
Mean = 640 ÷ 8 = 80			**Mean** = 640 ÷ 8 = 80		

Standard deviation =

$$\sqrt{\frac{\text{Sum of (deviations)}^2}{\text{Number of scores}}} = \sqrt{\frac{204}{8}} = 5.0$$

Standard deviation =

$$\sqrt{\frac{\text{Sum of (deviations)}^2}{\text{Number of scores}}} = \sqrt{\frac{2000}{8}} = 15.8$$

Most cases fall near the mean, and fewer cases fall near either extreme. This bell-shaped distribution is so typical that we call the curve it forms the **normal curve.**

As **FIGURE 7.3** shows, a useful property of the normal curve is that roughly 68 percent of the cases fall within one standard deviation on either side of the mean. About 95 percent of cases fall within two standard deviations. Thus, as Module 61 notes, about 68 percent of people taking an intelligence test will score within ±15 points of 100. About 95 percent will score within ±30 points.

normal curve *(normal distribution)* a symmetrical, bell-shaped curve that describes the distribution of many types of data; most scores fall near the mean (about 68 percent fall within one standard deviation of it) and fewer and fewer near the extremes.

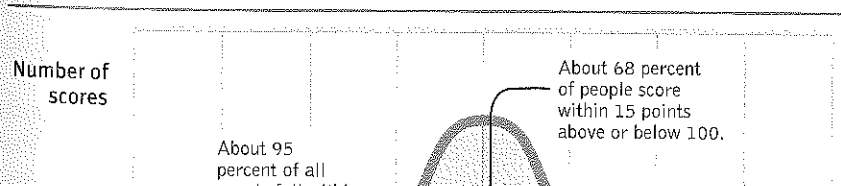


Figure 7.3

The normal curve Scores on aptitude tests tend to form a normal, or bell-shaped, curve. For example, the most commonly used intelligence test, the Wechsler Adult Intelligence Scale

Inferential Statistics

How do we know whether an observed difference can be generalized to other populations?

Data are "noisy." The average score in one group (breast-fed babies) could conceivably differ from the average score in another group (bottle-fed babies) not because of any real difference but merely because of chance fluctuations in the people sampled. How confidently, then, can we infer that an observed difference is not just a fluke—a chance result of your sampling? For guidance, we can ask how reliable and significant the differences are. These **inferential statistics** help us determine if results can be generalized to a larger population.

inferential statistics numerical data that allow one to generalize—to infer from sample data the probability of something being true of a population.

When Is an Observed Difference Reliable?

In deciding when it is safe to generalize from a sample, we should keep three principles in mind.

1. **Representative samples are better than biased samples.** As noted in Module 5, 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. Research never randomly samples 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 game-to-game points were consistent, an average is more reliable when it comes from scores with low variability.
3. **More cases are better than fewer.** An eager high school senior visits two university 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 (discounting the small sample size of only two instructors at each institution) tells friends about the "great instructors" at the first school, and the "bores" at the second. Again, we know it but we ignore it: *Averages based on many cases are more reliable* (less variable) than averages based on only a few cases.

The point to remember: Smart thinkers are not overly impressed by a few anecdotes. Generalizations based on a few unrepresentative cases are unreliable.

When Is a Difference Significant?

Perhaps you've compared men's and women's scores on a laboratory test of aggression, and found a gender difference. But individuals differ. How likely is it that the gender difference you found was just a fluke? Statistical testing can estimate the probability of the result occurring by chance.

Here is the underlying logic: When averages from two samples are each reliable measures of their respective populations (as when each is based on many observations that served difference can be generalized

(breast-fed babies) could conceivably dif-
le-fed babies) not because of any real dif-
s in the people sampled. How confidently,
not just a fluke—a chance result of your
and significant the differences are. These
can be generalized to a larger population.

e?

mple, we should keep three principles in

ed samples. As noted in Module 5, the
ptional and memorable cases one finds
le of cases. Research never randomly
pays to keep in mind what population

than those that are more variable. As
r whose game-to-game points were
omes from scores with low variability.
h school senior visits two university
t randomly attends two classes and
ing. At the next campus, the two
Returning home, the student
structors at each institution) tells
chool, and the "bores" at the second.
d on many cases are more reliable (less
.

verly impressed by a few anecdotes.
es are unreliable.

n a laboratory test of aggression, and
likely is it that the gender difference
nate the probability of the result oc-

two samples are each reliable mea-
s based on many observations that
reliable as well. (Example: The less

reasonable doubt means not making much of a finding unless the odds of its occurring by chance, if no real effect exists, are less than 5 percent.

When reading about research, you should remember that, given large enough samples, a difference between them may be "statistically significant" yet have little practical significance. For example, comparisons of intelligence test scores among hundreds of thousands of first-born and later-born individuals indicate a highly significant tendency for first-born individuals to have higher average scores than their later-born siblings (Kristensen & Bjerkedal, 2007; Zajonc & Markus, 1975). But because the scores differ by only one to three points, the difference has little practical importance.

The point to remember: Statistical significance indicates the *likelihood* that a result will happen by chance. But this does not say anything about the *importance* of the result.

AP® Exam Tip

Sometimes a phrase that is frequently used in the media has a more specific meaning when used in psychology. That's the case with the phrase "statistically significant." Make sure you know the precise meaning.

Before You Move On

▶ ASK YOURSELF

Find a graph in a popular magazine ad. How does the advertiser use (or abuse) statistics to make a point?

▶ TEST YOURSELF

Can you solve this puzzle?

The registrar's office at the University of Michigan has found that usually about 100 students in Arts and Sciences have perfect grades at the end of their first term at the University. However, only about 10 to 15 students graduate with perfect grades. What do you think is the most likely explanation for the fact that there are more perfect grades after one term than at graduation (Jepson et al., 1983)?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

Module 7 Review

How do we describe data using three mea-

7-2 How do we know whether an observed difference can be generalized to other populations?

- To feel confident about generalizing an observed difference to other populations, we would want to know that
  - the sample studied was representative of the larger population being studied;
  - the observations, on average, had low variability;
  - the sample consisted of more than a few cases; and
  - the observed difference was *statistically significant*.

Multiple-Choice Questions

1. Which of the following is a measure of variation?

a. Range
b. Mean
c. Mode
d. Frequency
e. Median

2. Which statistical measure of central tendency is most affected by extreme scores?

a. Mean
b. Median
c. Mode
d. Skew
e. Correlation

3. A researcher calculates statistical significance for her study and finds a 5 percent chance that results are due to chance. Which of the following is an accurate interpretation of this finding?

a. This is well beyond the range of statistical significance.
b. This is the minimum result typically considered statistically significant.
c. This is not statistically significant.
d. There is no way to determine statistical significance without replication of the study.
e. Chance or coincidence is unrelated to statistical

4. Descriptive statistics ________, while inferential statistics ________.

a. indicate the significance of the data; summarize the data
b. describe data from experiments; describe data from surveys and case studies
c. are measures of central tendency; are measures of variance
d. determine if data can be generalized to other populations; summarize data
e. summarize data; determine if data can be generalized to other populations

5. In a normal distribution, what percentage of the scores in the distribution falls within one standard deviation on either side of the mean?

a. 34 percent
b. 40 percent
c. 50 percent
d. 68 percent
e. 95 percent

on average, had low variability;
ted of more than a few cases; and
rence was *statistically significant.*

________, while inferential

ficance of the data; summarize the

n experiments; describe data from
studies
entral tendency; are measures of

can be generalized to other
marize data
determine if data can be generalized
ons

on, what percentage of the scores
s within one standard deviation on
n?

Practice FRQs

1. Explain the difference between descriptive and inferential statistics in research.

Answer (2 points)

1 point: Descriptive statistics organize and summarize the data collected during research.

1 point: Inferential statistics are used to help determine whether results can be generalized to a larger population through the calculation of statistical significance.

2. The following data set includes information from survey research in a psychology course regarding how many hours each individual in the class spent preparing for the exam.

Student	Amount of hours reported studying
1	2
2	3
3	6
4	8
5	9
6	9
7	21

Examine the data and respond to the following:

- What is the middle score in this distribution? What term is used to describe the middle score?
- What would be the most useful statistic for measuring the variation of the hours spent studying? Why is this statistic a better measure of variation than the range?

(3 points)