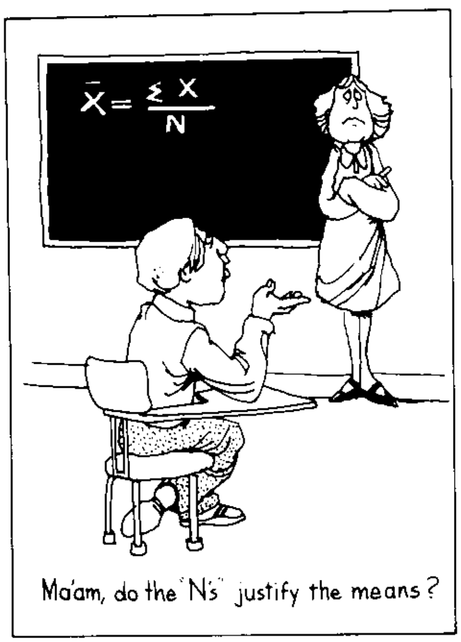
Whenever a graph is presented in which the base of the ordinate is not set at zero, be on the alert. The stage has been set for a sleight-of-hand trick. Without question, the most serious error in graphing is using a value other than zero as a base. Where there is no visual relationship to zero, the values are without perspective and thus have no meaning. It's like someone telling you that the temperature outside is 40°, not indicating the scale—Fahrenheit or Celsius—or whether it's 40° above or below zero. You might dress to prevent frostbite and end up suffering a heatstroke.

MEASURES OF CENTRAL TENDENCY

The tools called **measures of central tendency** are designed to give information concerning the average or typical score of a large number of scores. If, for example, we were presented with all the IQ scores of the students at a college, we could utilize the measures of central tendency to give us some description of the typical, or average, intellectual level at that school. There are three methods for obtaining a measure of the central tendency. When used appropriately, each is designed to give the most accurate estimation possible of the typical score. Choosing the appropriate method can sometimes be tricky. The interpretation of the data can vary widely, depending on how the typical score has been found.



The Mean

The most v
the **mean,**
is calculate
ber of scor

$\bar{X}$ =

In
stands for
Greek capi
all measure
N stands f
that you fo
something
the summa
In
been calcu

Table 2.3 (
t

X
130
120
115
115
110
110
105
100
100
100
95
95
90
85
75
ΣX = 1545

$\bar{X} = \frac{\Sigma X}{N}$ =

The Mean

The most widely, though not always correctly, used measure of central tendency is the **mean,** symbolized as $\bar{X}$. The mean is the arithmetic average of all the scores. It is calculated by adding all the scores together and then dividing by the total number of scores involved. The formula for the mean is as follows:

$$\bar{X} = \frac{\Sigma X}{N}$$

In the formula, $\bar{X}$, of course, stands for the mean. The capital letter X stands for the raw score, or the measure of the trait or concept in question. The Greek capital letter Σ (sigma) is an operational term that indicates the addition of all measures of X. This is usually read as "summation of." Finally, the capital letter N stands for the entire number of observations being dealt with. (It is important that you follow the book's use of capital letters, since lowercase letters often mean something quite different.) Thus, the equation tells us that the mean ($\bar{X}$) is equal to the summation (Σ) of all the raw scores (X) divided by the number of cases (N).

In Table 2.3, the mean of the distribution of IQ scores from Table 2.1 has been calculated. It happens that the mean is an appropriate measure of central

Table 2.3 Calculation of the mean from a distribution of raw scores.

	X
	130
	120
	115
	115
	110
	110
	105
	100
	100
	100
	95
	95
	90
	85
	75
$\Sigma X =$	1545

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1545}{15} = 103$$

Table 2.4 Calculation of the mean from a distribution of raw scores where the result is not a whole number.

X
72
71
70
68
68
68
65
63
$\Sigma X = 545$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{545}{8} = 68.125 = 68.13$$

tendency in this case because the distribution is fairly well balanced. Most of the scores occur in the middle range, and there are no extreme scores in either direction. Since the mean is calculated by adding together all of the scores in the distribution, it is not usually influenced by the presence of extreme scores, *unless* the extreme scores are all at one end of the range. The mean is typically a stable measure of central tendency, and, without question, it is the most widely used.

Calculating the Mean. Note in Table 2.3 that the mean is calculated as 103, a whole number. In most situations, however, this won't be the case. A more typical distribution is shown in Table 2.4. The mean in this case is rounded to 68.13.

*Always round to two places.** This requires completing the calculations to three places to the right of the decimal, then rounding the value that is two places to the right of the decimal. If the value in the third place is a 5 or higher, raise the value in the second place by 1. Thus, 68.125 becomes 68.13. But, 68.124 is rounded to 68.12. Remember, whenever you multiply or divide, square or take a square root, you must round your answer to two places. (This is accurate enough for a first course in statistics. When doing a research report for a class or for presentation, three- or four-place accuracy is often required.)

Interpreting the Mean. Interpreting the mean correctly can sometimes be a challenge, especially in situations where either the group or the size of the group changes. For example, the mean IQ of the typical freshman college class is about

*Although some statistics texts maintain that a value of 5 in the third place to the right of the decimal should always be rounded to the nearest even number, this book uses the convention of rounding a 5 to the next highest number. This is consistent with the rounding program built into most modern calculators. If your calculator has the fixed decimal feature, simply set it for two places, and the rounding will take place automatically.

115, whereas the mean IQ of the typical senior class is about 5 points higher. Does this indicate that students increase their IQs as they progress through college? No, but since the size of the senior class is almost always smaller than the size of the freshman class, the two populations are not the same. Among those freshmen who never become seniors are a goodly number with low IQs, and their scores are not being averaged in when their class later becomes seniors.

The Mean of Skewed Distributions. In some situations, the use of the mean can lead to an extremely distorted picture of the average value of a distribution of scores. For example, look at the distribution of annual incomes in Table 2.5. Note that one income ($10,000,000.00) is so extremely far above the others that the use of the mean income as a reflection of averageness gives a highly misleading picture of great prosperity for this group of citizens. A distribution like this one, which is unbalanced by an extreme score at or near one end, is said to be a **skewed distribution.**

Figure 2.8 shows what skewed distributions look like in graphic form. In the distribution on the left, most of the scores fall to the right, or at the high end, and there are only a few extremely low scores. This is called a negatively skewed, or skewed to the left, distribution. (The skew is in the direction of the tail-off of scores, not of the majority of scores.) The distribution on the right represents the opposite situation. Here most of the scores fall to the left, or at the low end of the distribution, and only a very few scores are high. This is a positively skewed, or skewed to the right, distribution. Remember, label skewed distributions accord-

Table 2.5 Calculation of the mean of a distribution of annual incomes.

	X
	$10,000,000.00
	20,000.00
	20,000.00
	19,500.00
	19,400.00
	19,400.00
	19,400.00
	19,300.00
	19,000.00
	18,500.00
	18,000.00
	18,000.00
	17,600.00
$\Sigma X =$	$10,228,100.00

$$\bar{X} = \frac{\Sigma X}{N} = \frac{10{,}228{,}100.00}{13} = \$786{,}776.92$$

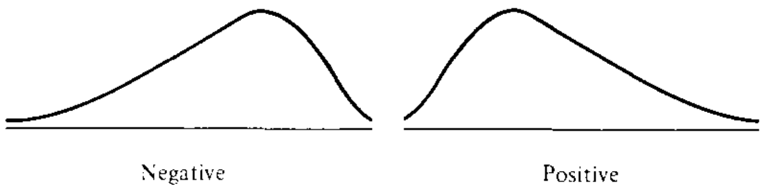


Figure 2.8 A graphic illustration of skewed distributions.

ing to the direction of the tail. When the tail goes to the left, the curve is negatively skewed; when it goes to the right, it is positively skewed.

Thanks, Ralph. A rather dramatic example of how the use of the mean distorts "averageness" in a skewed distribution can be seen in a University of Virginia press release, undoubtedly meant as tongue-in-cheek. During an analysis of how well the members of the college's 1983 graduating class fared in the job market, it was discovered that the highest salaries were earned by graduates of the Department of Rhetoric and Communications Studies, where the beginning average pay was $55,000 a year. It should be pointed out that one student, the 7-foot 4-inch-tall basketball player Ralph Sampson, had a starting salary of well over $1 million. Perhaps the mean height for those same graduates was 6 feet 6 inches.

The Median

The **median** is the exact midpoint of any distribution, or the point that separates the upper half from the lower half. In fact, the median (symbolized as Mdn) is a much more accurate representation of central tendency for a skewed distribution than is the mean. Whereas the mean income of the distribution in Table 2.5 is $786,776.92, the median income is $19,400.00, a much more descriptive reflection of the typical income for this distribution. Since income distributions are almost always skewed toward the high side, you should be on the alert for an inflated figure whenever the mean income is reported. The median gives a better estimation of how the typical wage earner is actually faring.

As a memory trick for recalling the definition of the median, think of the median strip dividing a highway. The same width of road lies both to the left of the median strip and to the right.

Calculating the Median. To calculate the median, the scores must first be arranged in distribution form, that is, in order of magnitude. Then, count down (or up) through half of the scores. For example, in Table 2.5 there are 13 income scores in the distribution. Therefore, count down 6 scores from the top, and the seventh score is the median (there will be the same number of scores above the median point as there are below it). Whenever a distribution contains an odd number of scores, finding the median is very simple. Also, in such distributions

Table 2.6

X
120
118
115
114
114
112
$\Sigma X = 693$
$\bar{X} = 115.$

the medi
tion in T
I
slightly d
The med
two midd
fact that
ceived th

T
affected
direction
even wit
mean of
value of
distributi
reflect tr
larly too

Table 2.7

X
120
118
115
114
114
6
$\Sigma X = 587$
$\bar{X} = 97.8$

Table 2.6 Calculation of the median with an even number of scores.

X	
120	
118	
115	— 114.5 Median
114	
114	
112	
$\Sigma X = 693$	
$\bar{X} = 115.50$ Mean	

the median will usually be a score that someone actually received. In the distribution in Table 2.5, someone really did earn the median income of $19,400.

If a distribution is made up of an even number of scores, the procedure is slightly different. Table 2.6 presents a distribution of an even number of scores. The median is then found by determining the score that lies halfway between the two middlemost scores. In this case, the median is 114.5. Don't be disturbed by the fact that in some distributions of even numbers of scores nobody actually received the median score; after all, nobody ever had 2.8 children either.

The Median of Skewed Distributions. Unlike the mean, the median is not affected by skewed distributions (where there are a few extreme scores in one direction). In Table 2.7, for example, the median score is still found to be 114.5, even with a low score of 6 rather than one of 112 as reported in Table 2.6. The mean of this distribution, on the other hand, plummets to an unrepresentative value of 97.83. The mean is always pulled toward the extreme score in a skewed distribution. When the extreme score is at the high end, the mean is too high to reflect true centrality; when the extreme score is at the low end, the mean is similarly too low.

Table 2.7 Calculation of the median with an even number of scores and a skewed distribution.

X	
120	
118	
115	— 114.5 Median
114	
114	
6	
$\Sigma X = 587$	
$\bar{X} = 97.83$ Mean	

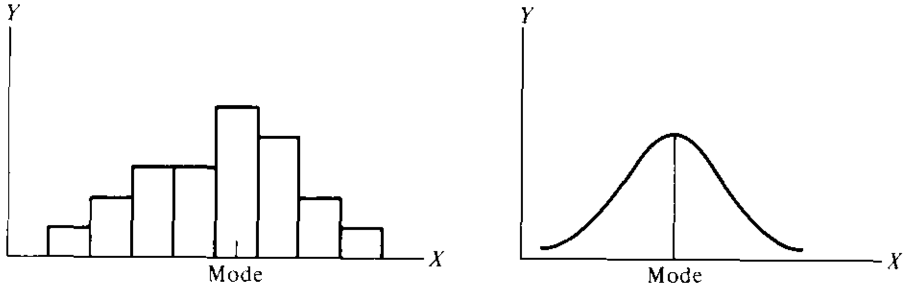


Figure 2.9 The location of the mode in a histogram (left) and a frequency polygon (right).

The Mode

The third, and final, measure of central tendency is called the **mode** and is symbolized as Mo. The mode is the most popular, or most frequently occurring, score in a distribution. In a histogram, the mode is always located beneath the tallest bar; in a frequency polygon, the mode is always found directly below the point where the curve is at its highest. This is because, as was pointed out previously, the Y axis, or the ordinate, represents the frequency of occurrence. (See Fig. 2.9.)

Finding the Mode. When the data are not graphed, just determine which score occurs the most times, and you've got the mode. For example, in the distribution shown in Table 2.8, the score of 103 occurs more often than does any other score. That value is the mode. In Table 2.9 a frequency distribution of the same data is given. Here, to find the mode, all you have to do is to note which score (X) is beside the highest frequency value (f). The mode is a handy tool since it pro-

Table 2.8 Finding the mode of a distribution of raw scores.

X	
110	
105	
105	
103	Mode
103	Mode
103	Mode
103	Mode
101	
101	
100	
100	
98	
95	

Table 2.9 Finding the mode when a frequency distribution is given.

X	f
110	1
105	2
103 ← Mode	4
101	2
100	2
98	1
95	1

vides an extremely quick method for obtaining some idea of a distribution's centrality. Just eyeballing the data is usually enough to spot the mode.

Bimodal Distributions. A distribution having a single mode is referred to as a **unimodal distribution.** However, some distributions have more than one mode. When there are two modes, as in Fig. 2.10, the distribution is called bimodal. (When there are more than two modes, it is called multimodal.) Distributions of this type occur when scores cluster together at several points, or if the group being measured really represents two or more subgroups.

Assume that the distribution in Fig. 2.10 represents the running times (in seconds) in the 100-yard dash for a large group of high school seniors. There are two modes—one at 13 seconds and the other at 18 seconds. Since there are two scores that both occur with the same high frequency, it is probable that data about two separate subgroups are being displayed. For example, perhaps the running times for boys are clustering around one mode and the speeds for girls are clustering around the other.

Interpreting the Bimodal Distributions. Whenever a distribution does fall into two distinct clusters of scores, extreme care must be taken with their interpretation. Neither the mean nor the median can justifiably be used, since a bimodal distribution cannot be adequately described with a single value. A person whose

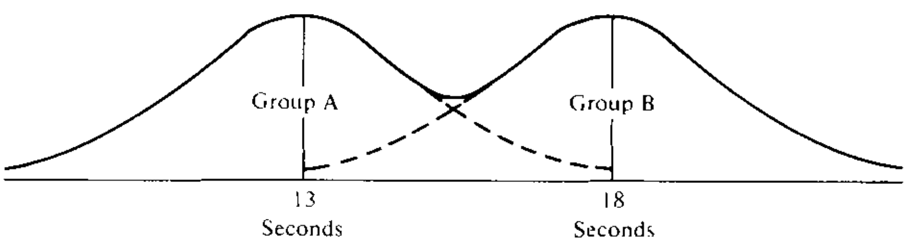


Figure 2.10 Graph of bimodal distribution of running times.

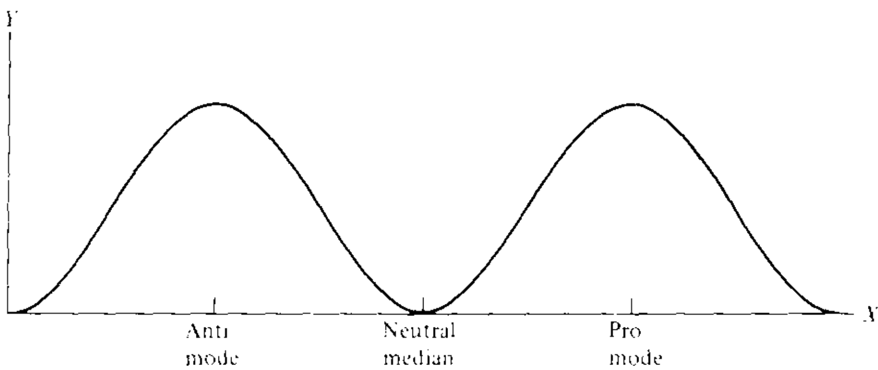


Figure 2.11 Graph of responses to an attitude questionnaire showing two modes.

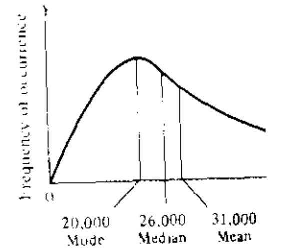
head is packed in ice and whose feet are sitting in a tub of boiling water cannot be appropriately characterized as being in a tepid condition *on the average*.

Bimodal distributions should not be represented by the use of a single average of the scores. Suppose a group of individuals completed a certain attitude questionnaire. The resulting scores fell in two decidedly different clusters, half the group scoring around a "pro" attitude and the other half around an "anti" attitude. The use of either the mean or the median to report the results would provide a highly misleading interpretation of the group's performance, for in either case the group's attitude would be represented as being neutral. Figure 2.11 shows such a bimodal distribution. The two modes clearly indicate how divided the group really was. Note, too, that while *nobody* scored at the neutral point, using either the mean or the median as a description of centrality would imply that the typical individual in the group was indeed neutral. Thus, when a distribution has more than one mode, the modes themselves, not the mean or the median, should be used to provide an accurate account.

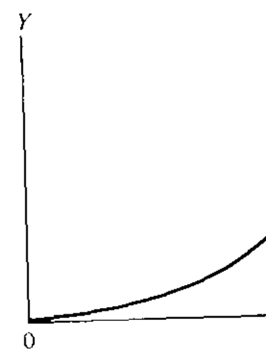
APPROPRIATE USE OF THE MEAN, THE MEDIAN AND THE MODE

Working with Skewed Distributions

The best way to illustrate the comparative applicability of the three measures of central tendency is to look again at a skewed distribution. Figure 2.12 shows an approximation of the income distribution per household in the United States in 1987. Like most income distributions, this one is skewed to the right. This is because the low end has a fixed limit of zero, while the sky is the limit at the high end. Note that the exact midpoint of the distribution, the median, falls at a value of $26,000 a year. This is the figure above which 50% of the incomes fall and below which 50% fall. Because there is a positive skew, the mean indicates a fairly



high average i
ture of reality,
high end of th
which is $20,0
the mode doe
lower than the
skewed distrib
ent, portraits
tively skewed
from left to ri
point; and fin
simply reverse
tributions is tl
A mn
you go up the
alphabetical o
this memory
you're going
area, riding th



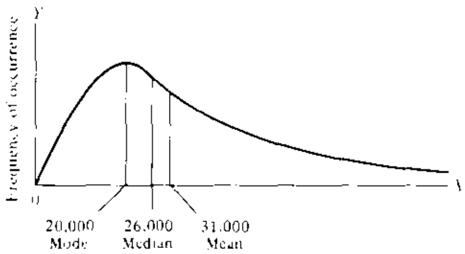


Figure 2.12
Distribution of income per household in the United States

high average income of $31,000. This value, however, gives a rather distorted picture of reality, since the mean is being unduly influenced by the few families at the high end of the curve whose income is in the millions. Finally, the modal income, which is $20,000 per year, seems to distort reality toward the low side. Although the mode does represent the most frequently earned income, it is nevertheless lower than the point separating the income scores into two halves. In the case of a skewed distribution, then, both the mean and the mode give false, though different, portraits of typicality. The truth lies somewhere in between. Thus, in a positively skewed distribution the order of the three measures of central tendency from left to right is first the mode, the lowest value; then the median, the midpoint; and finally the mean, the highest value. A negatively skewed distribution simply reverses this order. (See Fig. 2.13). The point to remember for skewed distributions is that the mean is always located toward the tail end of the curve.

A mnemonic device often used for this purpose is to remember that as you go up the slope of a skewed curve, the measures of central tendency appear in alphabetical order, first the mean, then the median, and finally the mode. In using this memory prod, you must remember that you're going *up* the slope and that you're going up on the gentle slope, *not* the steep side. Picture yourself at a ski area, riding the chairlift up the novice slope.

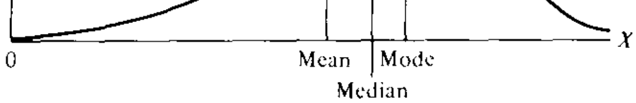
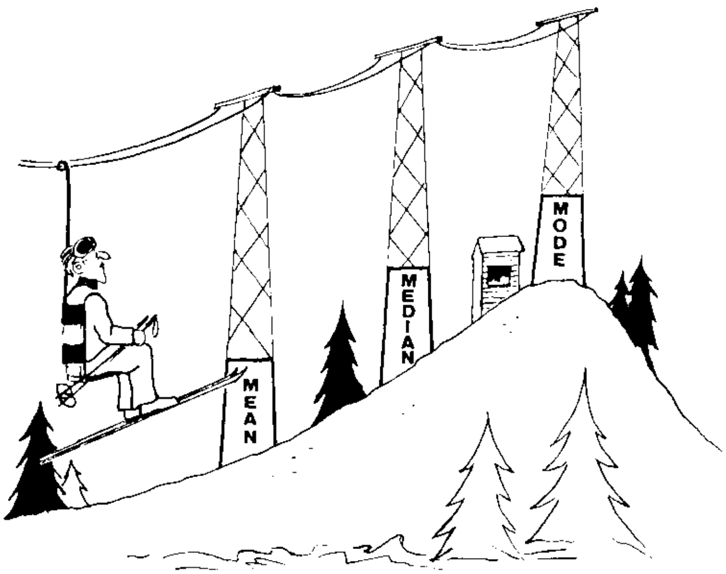


Figure 2.13
A negatively skewed distribution



Assessing Skewness (Sk)

Although there are techniques for establishing a precise value for the amount of skewness (Sk) present in any unimodal distribution of scores (see the Appendix), a somewhat rough but easy method for spotting Sk is simply to compare the mean and median. Remembering that in skewed distributions the mean always "tips toward the tail," then the mean can be used as a quick indicator of skewness. If the mean lies to the left of the median, the distribution is skewed left. If it lies to the right of the median, the distribution is skewed to the right. Also, the greater the distance between the mean and median, the greater the total amount of Sk. For example, if you are told that the mean price for a new home in the United States is $100,000 and that the median price is $80,000, then you may conclude that the distribution of prices is skewed to the right (which it really is), or has an Sk+. Or, if the median golf score for the members of a certain country club happened to be 95, but the mean was only 85, then the distribution of golf scores would show a negative skew, or Sk−. In general, the greater the discrepancy between the mean and the median, the greater the skewness.

By incorrectly using such descriptive statistics as the measures of central tendency, some extremely interesting bits of con-artistry can be performed. As previously mentioned, if a distribution is skewed to the right, the mode reflects the lowest value of central tendency while the mean portrays the highest. With such a distribution, a researcher could report on a before-and-after study and make it appear that group performance had increased in the "after" condition, even though *every score in the entire group had remained the same.* How? The researcher simply reports the modal score for the results of the "before" condition and the

mean score f
measure of ce

Effects of the

Which of the
measurement
has assumed
cessive scale
measures of c

Ordi
nal data prov
calculating a
to the assum
can and shou
tion, it is itse
half the scor
scores (arran

Final
since each o
than. Nomin
form of freq
measure of c
frequently occ
occurrence,
to handle.

In sh
central tend
anced, use t
median. If th
ent scale pos
applicable,
ferent ordin
choice but t

Late
techniques,
dency will b

SUMMARY

Descriptive
of data in c

mean score for the results of the "after" condition. Does it matter, then, which measure of central tendency is used? You bet it does!

Effects of the Scale of Measurement Used

Which of the measures of central tendency to use also depends on the scale of measurement on which the data are based. The discussion so far in this chapter has assumed an interval scale, that is, a scale in which the difference between successive scale points is equal. Interval data allow for the calculation of all three measures of central tendency—mean, median, and mode.

Ordinal data, however, cannot be used to calculate the mean. Since ordinal data provide no information regarding the distance between the scale points, calculating an ordinal mean can be extremely misleading. Such a mean gives rise to the assumption that the distances are known. With ordinal data, the median can and should be used. Since the median is the middlemost point in a distribution, it is itself a rank—a rank above which half the scores fall and below which half the scores fall. Recall that before finding the median you had to rank the scores (arrange them in order of magnitude).

Finally, with nominal data, neither the mean nor the median can be used, since each of these measures implies the comparisons of greater than and less than. Nominal data are restricted by the equality-nonequality rule and are in the form of frequency of occurrence within discrete categories. Therefore, the only measure of central tendency permissible for nominal data is the mode, the *most frequently occurring* score. The mode is determined on the basis of frequency of occurrence, which is precisely the kind of data that the nominal scale is designed to handle.

In short, interval data and ratio data allow the use of all three measures of central tendency. If the interval data distribution is unimodal and fairly well balanced, use the mean. If the distribution is skewed to the right or the left, use the median. If the distribution has more than one clustering of score values at different scale positions, use the mode. With ordinal data, since the mean is no longer applicable, use the median. However, if there are clusterings of tied ranks at different ordinal positions, then use the mode. If the data are nominal, there is no choice but to use the mode.

Later in this book, during the discussion of nonparametric statistical techniques, other more subtle distinctions among the measures of central tendency will be presented.

SUMMARY

Descriptive statistics includes those techniques used for describing large amounts of data in convenient and symbolic form.