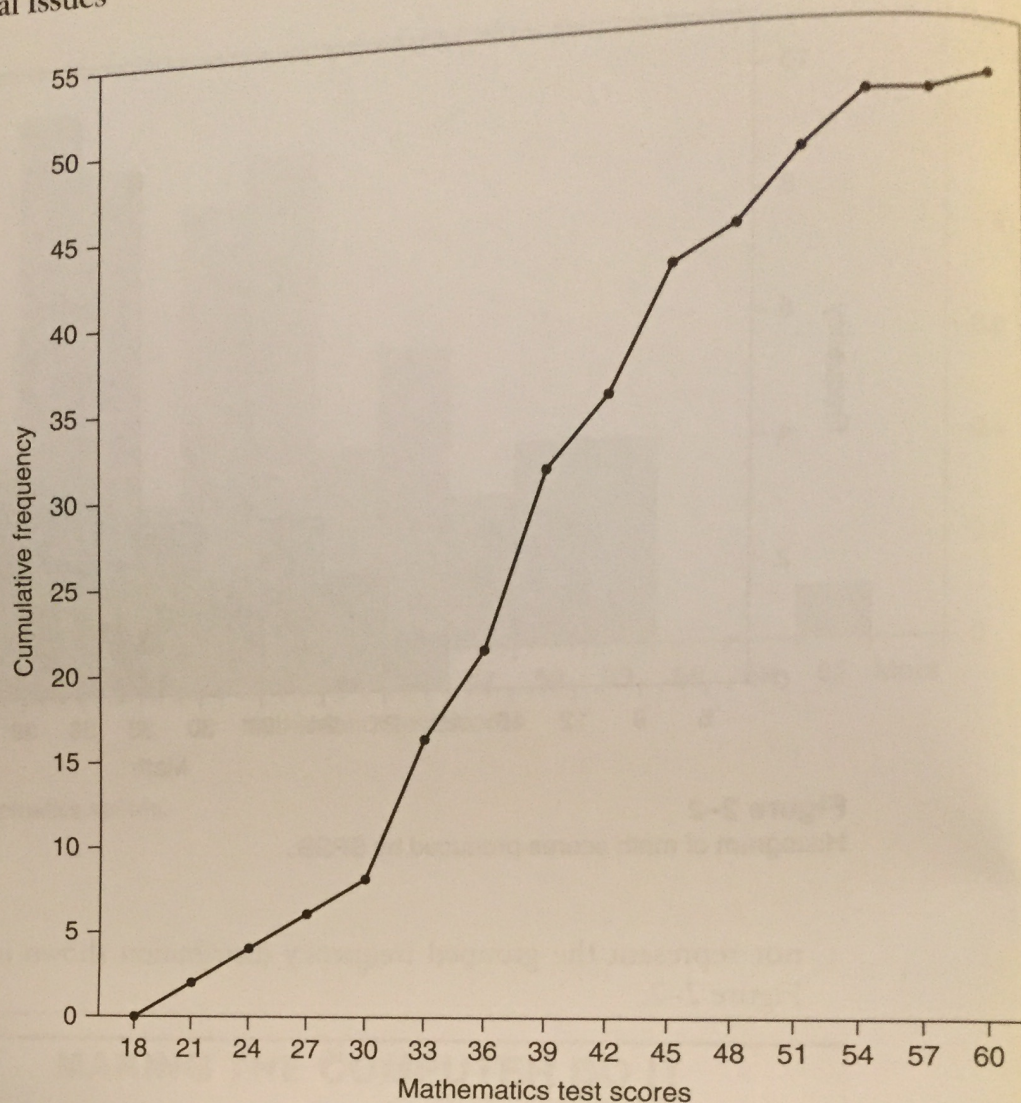


Figure 2-3
Cumulative frequency curve.



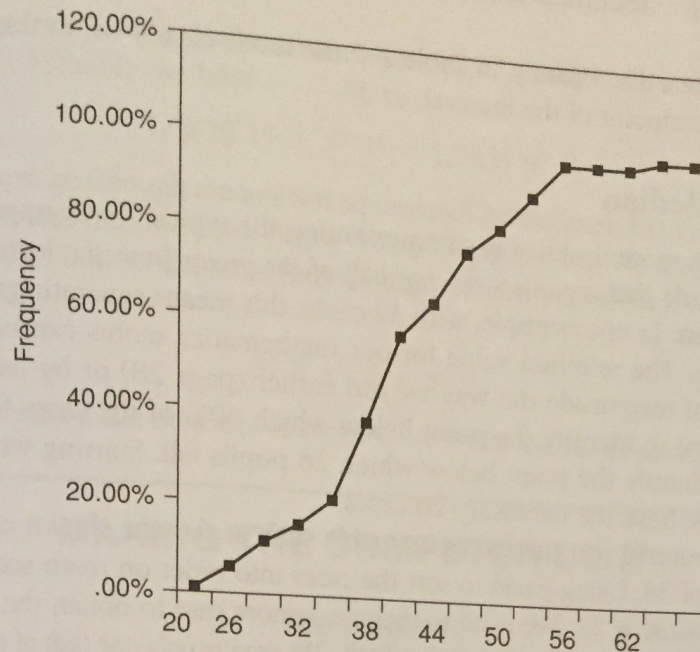
MAKING THE COMPUTER DO IT

Cumulative Frequency Curves

Excel will prepare a graph of the cumulative percentage distribution if you select both Cumulative Percentage and Chart Output. You will have to edit the chart to remove the extra space as described earlier, but the graph will correctly represent the cumulative frequencies from the **grouped frequency distribution**. Unfortunately, the ordinate will be represented as a percentage of the sample size rather than a number of cases. By deleting the histogram bars and rescaling the graph, you should get a figure like the one on page 39.

To obtain a graph of cumulative frequency or cumulative percentage from SPSS, click on the Graphs menu and select Line. Click on Define, select the variable you wish to graph, and place it in the Category Axis window. Select "Cum n of cases," then click OK. You can also graph the cumulative percentages by selecting "Cum% of cases." Unfortunately, the program produces a cumulative frequency curve from the program's cumulative frequency distribution. This means that the graph will omit values of the variable that have zero frequency in the same way that the frequency distribution itself does and will list all

Figure 2-4
Cumulative frequency curve produced by EXCEL.



distinct values that occur. Again, this requires caution in interpreting the graph because there is no warning that the program has omitted some values.

MEASURES OF CENTRAL TENDENCY

We often need a statistic to represent the typical, or average or middle, score of a group of scores. The object of these statistics is to provide a single measure to locate the distribution of scores on the score scale. We present three such statistics, each of which defines the center in a different way and conveys slightly different information.

The Mode

A very simple way of identifying the typical score is to pick the one that occurs most frequently. This score is called the **mode** and corresponds to the highest point in the histogram. If you examine the array of math scores in the tabulation in the Preparation of a Frequency Distribution section, you will find that the scores 33 and 38 each occur five times. A unique value for the mode does not exist with these data. If one of the students who scored 37 had instead scored 38, the mode would be 38. If one of the 34s had been a 33, then 33 would have been the mode. The mode is sensitive to such minor changes in the data and is therefore a crude and often not very useful indicator of the typical score. In Table 2-3, the grouped frequency distribution, the **modal interval** is 36-38. When the scores are grouped in this way, we can call the midpoint of the modal interval, 37, the *mode*.

Our two computer programs approach determination of the mode in the same way. Both look for the score value in the raw data that has the greatest frequency. When, as is often the case, two or more scores occur with the same frequency, the smaller value is reported as the mode. Thus, both programs report the mode for our set of math test scores as 33, even though the score 38 also had a frequency of five. SPSS reports that there are multiple modes, but Excel does not. When we compare this result with the highest point in our grouped frequency distribution, we

get quite a discrepancy. In Table 2-3 the mode clearly fell in the interval 36-38, yielding a mode at the midpoint of the interval, or 37.

The Median

A much more useful way of representing the typical, or average, score is to find *the value on the score scale that separates the top half of the group from the bottom half*. This value is called the **median**. In our example, with 52 cases, this means separating the top 26 students from the bottom 26. The required value for our mathematics scores can be found by placing the scores in order of magnitude the way we did earlier (page 28) or by using the Sort command in Excel. We want to identify the point below which 50% of the cases fall. Because 50% of 52 is 26, we must identify the point below which 26 pupils fall. Starting with the lowest score, we count up until we have the necessary 26 cases.

Counting just the scores in a table such as the one shown earlier, the 26th score is one of the scores of 38. Using Excel to sort the cases into order on math score, the 25th case is Gail Galaraga, whose score is 37. We need to include 1 more case to obtain the required 26 cases. The next score value (38) is shared by five individuals. *We require only one fifth of these individuals*. Now how shall we think of these cases being spread out over the score value of 38? As noted earlier in this chapter, a reasonable assumption is that they are spread out evenly over the interval. Then to include one fifth of the scores, we would have to go one fifth of the way from the bottom of the interval toward the top.

At this point, we must define what we mean by a score of 38. First, let us note that although test scores go by jumps (or discrete increments) of one unit (37, 38, and 39), we consider the underlying ability that the test measures to have a continuous distribution that takes in all the intermediate values between two scores. We might liken the situation to a digital clock. Although time is continuous, the recording instrument runs by jumps, with one jump every time the basic unit of 1 minute is passed.

Figure 2-5 illustrates this point. The bottom line represents the continuum of ability. We define 38 as the interval on the continuum that is closer to point 38 than to either 37 or 39. Thus, in Figure 2-5, 38 is represented as a slice extending from 37.5 to 38.5. Although somewhat arbitrary, this definition of a score is a reasonable one and is accepted by most authorities. The score interval 36-38 is really to be thought of as extending from 35.5 to 38.5 on the underlying continuum, as is shown in the figure. We do not get scores lying *between* 37 and 38 or between 38 and 39—not because those levels of the trait do not exist, but because our measuring instrument does not register any values between 37 and 38 or between 38 and 39. (This is not the way SPSS represents scores. In SPSS a score of 38 is read as extending from just above 37 up to exactly 38 and no higher. Thus, a score is at the upper limit of its interval rather than at the midpoint. The effect is usually small, but explains why you might get slightly different results from SPSS than you would from hand calculations or from another program.)

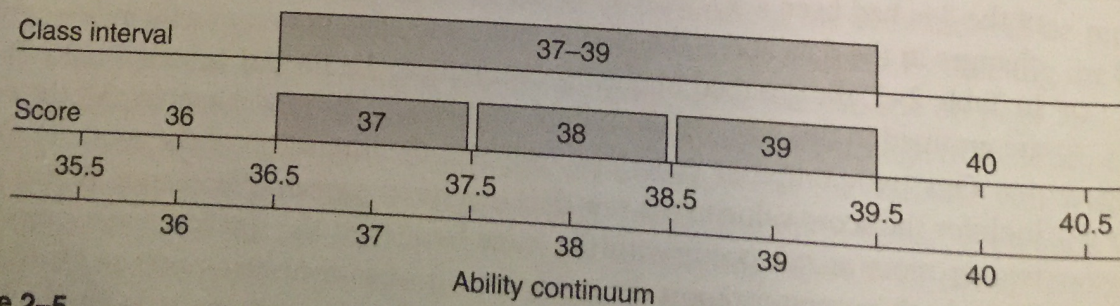


Figure 2-5
Relation between scores and ability continuum.

Because we require one fifth of the cases in interval 37.5–38.5, we must go one fifth of the way from 37.5 to 38.5; that is, we have

$$1/5(38.15 - 37.5) = 1/5(1) = .2$$

This is how far we must go through the interval represented by the score 38 to include one of the five people in the interval and find the point below which 26 people fall. We must add .2 to the value 37.5, which is the lower limit of the score interval that contains the median. Therefore, the median for this set of scores is

$$37.5 + .2 = 37.7$$

Note that the median need not be a whole score value while the mode always is.

MAKING THE COMPUTER DO IT

The Median

Most computer programs, including SPSS and Excel, define the median as the score obtained by the middle person, regardless of how many people got the same score. The programs look at the cumulative percent of cases for each score interval and select as the median the score corresponding to the first interval where the cumulative percent exceeds 50%. In our example, the interval for a score of 38 is the first interval with a cumulative percentage above 50 (it is actually 57.7). Both programs report 38 as the median. Because the differences found in computing the median by this approach and the one described above are usually small, they are not likely to be important to test users. However, when working with grouped data, the differences can be substantial. SPSS has a way to handle grouped data and produce results essentially identical to what we found here, but Excel does not. In the special case where the median falls exactly between two individuals who have different scores, the proper course to take is to find the point halfway between the two people's scores. For example, if we have 100 individuals where the 50th person's score is 30 and the 51st person's score is 32, the appropriate value for the median is halfway between the scores, or 31. Both Excel and SPSS will produce this result.

Percentiles

The same steps used to compute the median can be used to find the score below which any other percentage of the group falls. These values are called **percentiles**. The median is the 50th percentile, that is, the point on the score scale below which 50% of the individuals fall. If we want to find the 25th percentile, we must find the point on the score scale below which 25% of the cases fall; 25% of 52 is 13. For our set of 52 mathematics scores, 13 cases take us through the score value 32 ($cf = 11$), and include two of the five cases with a score of 33. So, the 25th percentile is located $2/5$ ths of the way through the interval 32.5–33.5 and is computed to be $32.5 + (2/5) = 32.9$.

As another illustration, consider the 85th percentile. We have $(.85)(52) = 44.2$ people making up the bottom 85% of our group. Because 44 cases carry us to the top of interval 47.5–48.5, and there are 2 cases in the next interval, for the 85th percentile, we need $.2/2 = .1$ of the interval and the 85th percentile is $48.5 + .1 = 48.6$. Other percentiles can be found in the same way.

To find the median or other percentiles from a grouped frequency distribution, we proceed in exactly the same way except that we must remember that our intervals are now more than one score unit wide. To find the median from the grouped frequency distribution in Table 2-3 we still need 26 cases. The 26th person is one of the 10 people in the interval 36-38. We have 20 people whose scores fall below this interval (*cf* of interval 33-35 is 20), so we need 6 of the 10 people in the interval to make our 26. Therefore, the median falls 6/10th of the way through the interval. Because the interval is 3 score units wide, we must go $(.6)(3) = 1.8$ score units (60%) into the interval to include the six people we need. The interval starts at 35.5 (the lower edge of the first score in the interval), so the median is

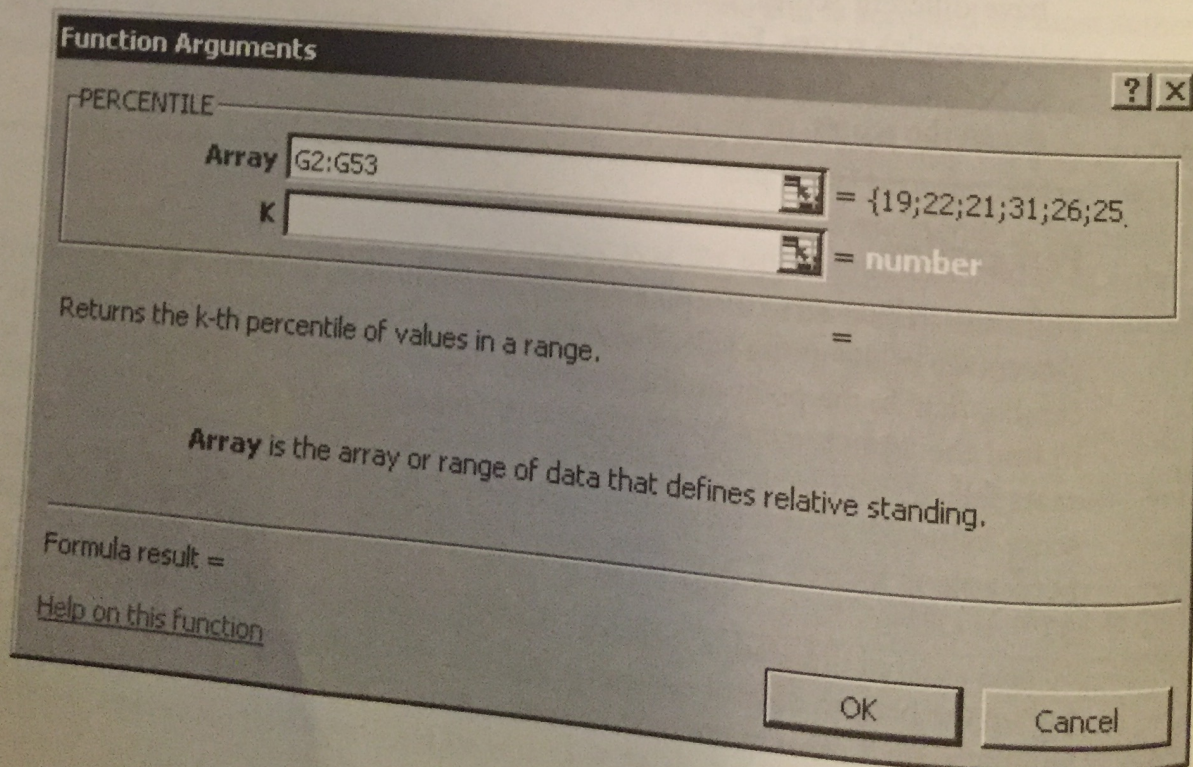
$$35.5 + 1.8 = 37.3$$

Percentiles have many uses, especially in connection with test norms and the interpretation of scores. We will encounter them again in Chapter 3.

MAKING THE COMPUTER DO IT

Percentiles

Excel computes percentiles using a different program function than we have discussed so far. On the Insert menu (and also as a button on the toolbar next to the sort icons) you can place a function (f_x) in a cell of the spreadsheet. First select a cell in which you wish the percentile to appear. Then click on the f_x symbol (some editions of Excel have an arrow next to the sum symbol Σ) and a series of function types such as "Math and Trig" and "Statistical" will appear. If you select the Statistical group, you will be presented with a wide variety of options, one of which is "Percentile." You can use this approach to compute any percentile you wish and most of the other descriptive statistics we will be discussing. After selecting Statistical, select Percentile and you will see a screen like this:



Select the cells containing the data you wish to include in the analysis (here the scores for our 52 students' math scores are in cells G2 to G53). Then click in the **K** cell and enter the percentile you wish (.50 for the median). Click OK and the desired percentile will appear in the selected cell.

Excel uses the same approach to find percentiles that it uses to find the median. That is, the program compares the specified percentile with the array of cumulative percentages and reports as the p th percentile (for example, the 17th) the first score value where the cumulative percentage is greater than p . The result is not as precise as the one described on page 42 (the same score may be reported as corresponding to several different percentiles), but it is usually accurate enough. We describe how to compute percentiles with SPSS in the next section.

The Arithmetic Mean

Another frequently used statistic for representing the middle of a group is the familiar average of everyday experience. Because statisticians speak of many measures of central tendency as averages, they identify this one as the **arithmetic mean (M)**. It is computed as the sum of a set of scores divided by the total number of scores. Thus, the arithmetic mean of the scores 4, 6, and 7 is

$$(4 + 6 + 7)/3 = 17/3 = 5.67$$

In our example of scores on the mathematics test, we can add the scores of all 52 individuals in the group, giving us 1,985. Dividing by 52, we get $M = 38.17$ for the average, or arithmetic mean, for this group.

We can express the process for computing the mean using a simple formula. Statisticians use the capital Greek letter sigma (Σ) to stand for the process of summation. If we use the letter X to stand for a variable, such as math test scores, then the expression ΣX tells us to sum the values of X . Because the mean requires that we divide this sum by the number of scores (N), an expression for the mean is

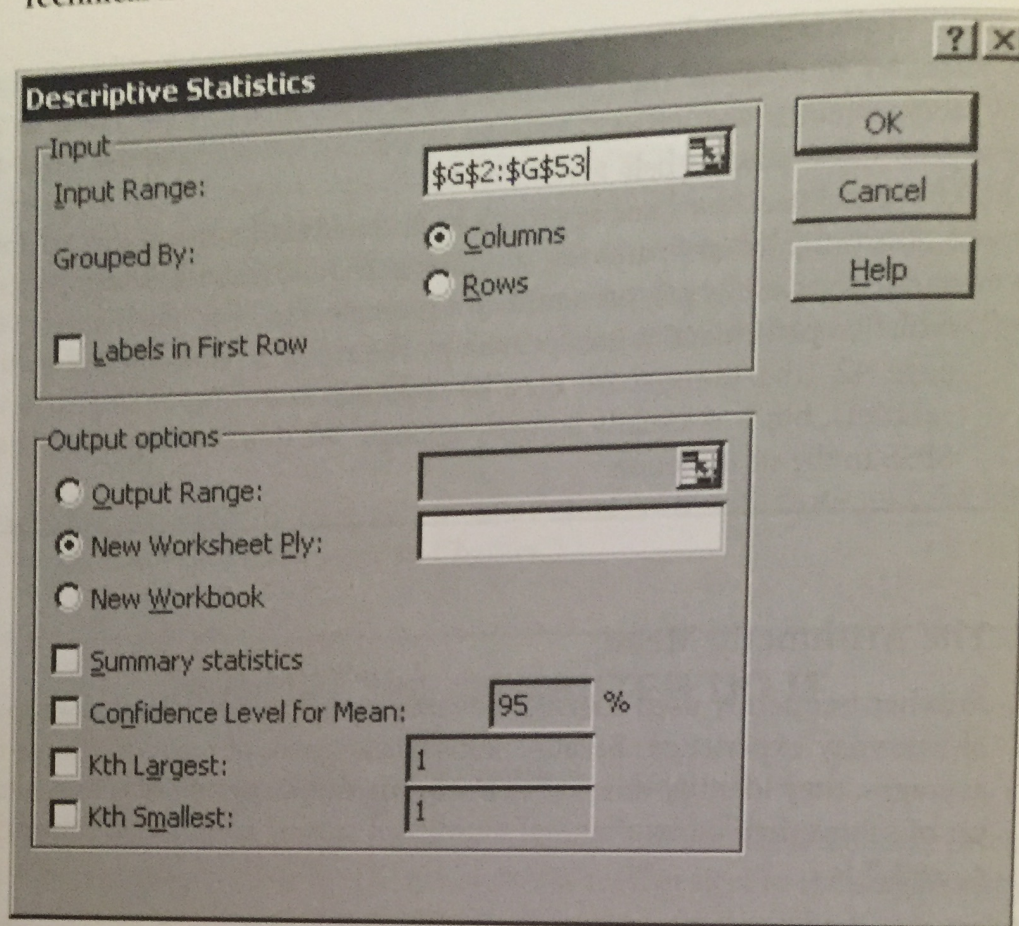
$$M = \frac{\Sigma X}{N} = \frac{1,985}{52} = 38.17$$

We will have many occasions to use this formula and other similar ones.

MAKING THE COMPUTER DO IT

The Arithmetic Mean

Excel and SPSS both have programs to compute the mean and several other measures that describe distributions of scores. In Excel, there is a program in the Data Analysis package called "Descriptive Statistics." Click on that program, then click OK and you will see a screen like the one on the next page.



Descriptive Statistics

Input
 Input Range: 

Grouped By: ☒ Columns ☐ Rows

☐ Labels in First Row

Output options
☐ Output Range:
☒ New Worksheet Ply:
☐ New Workbook

☐ Summary statistics

☐ Confidence Level for Mean: %

☐ Kth Largest:

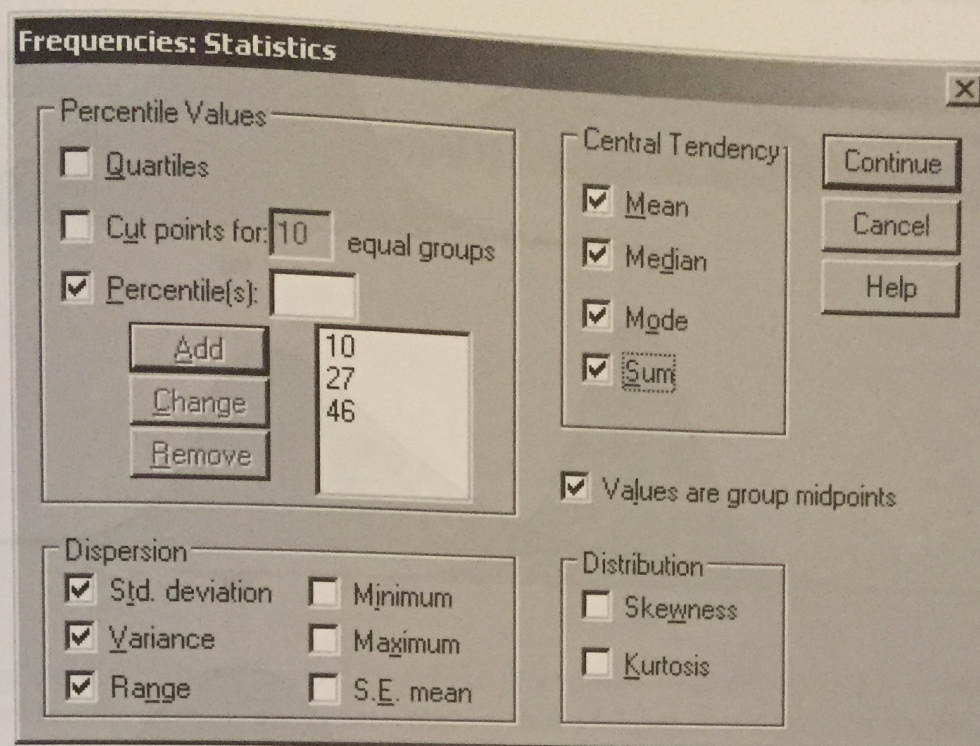
☐ Kth Smallest:

OK
 Cancel
 Help

Click in the Input Range window on the pop-up box, then on the main screen highlight the scores to be included in computing the mean. If you are using the data from Table 2-1, the math test scores are in column G, rows 2-53 as we show here. Next click in the "Output Range" window of a pop-up box. Then click on a cell in the spreadsheet where you would like the output to start, such as column I, row 1. Finally, make sure there is a check in the "Summary statistics" box, then click OK. The program will produce the three measures of central tendency we have described, plus 10 other statistics, some of which we will discuss shortly. Make sure you look at the "count" statistic to be sure you have included the correct number of scores in your computations. You can also use the statistical function (f) method (see the preceding box on percentiles) to obtain just the arithmetic mean. It is listed as the Average function.

If you are using SPSS to compute your descriptive statistics, you can get everything you need at the same time that you are preparing your frequency distribution. After opening the Frequencies program and selecting the variables you wish to analyze, click on the Statistics button to obtain the screen shown on the next page.

Click in the boxes next to the statistics you wish to compute. Here we have selected the mean, median, mode, and sum; three particular percentiles (10, 27, and 46); and three measures of variability to be described shortly. Placing a check mark in the "Values are group midpoints" box will cause SPSS to compute the median and percentiles as we have described them in this chapter. Failing to check that box will cause the program to take the first score value whose cumulative percentage exceeds the specified value. Click Continue, then OK to obtain your summary statistics.



Central Tendency and the Shape of the Distribution

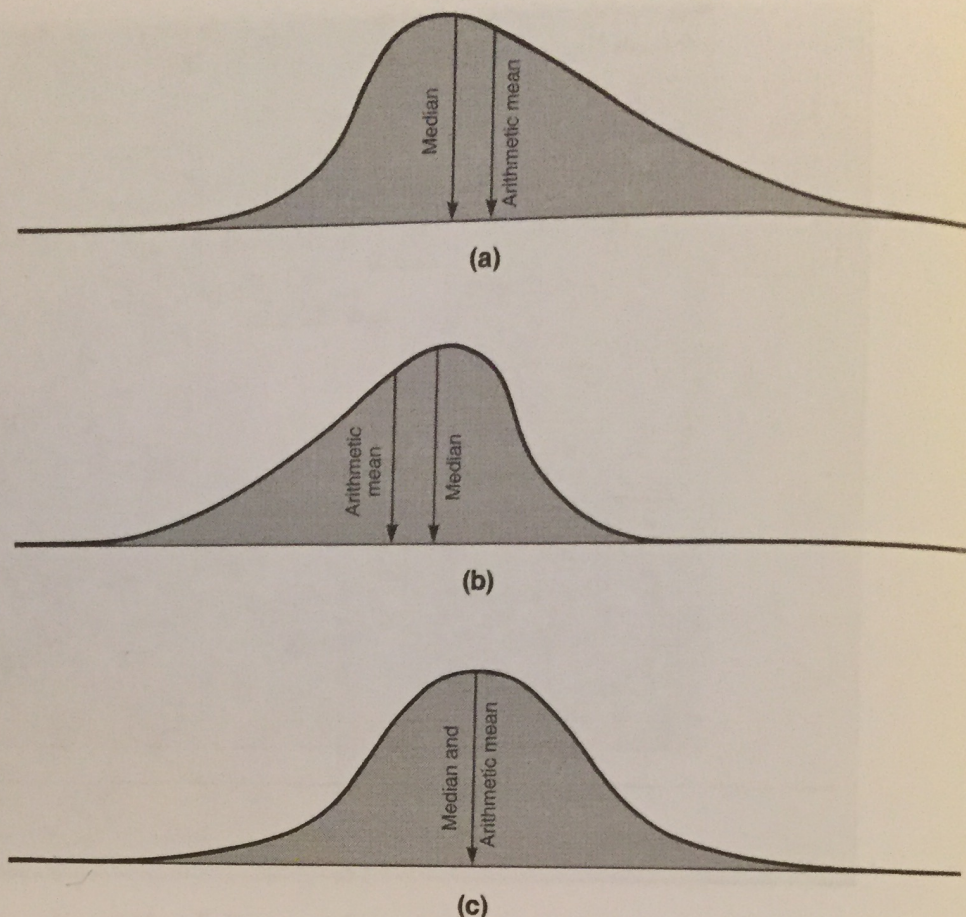
The mode, arithmetic mean, and median seldom have exactly the same value, but usually they do not differ greatly. In our example, the values of the median and mean are 37.7 and 38.17, respectively, and one alternative for the mode is 38. The three statistics will differ substantially only when the set of scores is *skewed* greatly, that is, when there is a piling up of scores at one end and a long, thin tail at the other. Figure 2–6 on page 46 shows three distributions that differ in the amount and direction of skewness. The top figure is positively skewed; that is, it has a tail running up to the high scores. We might get a distribution like this for income in the United States, because there are many people with small and moderate incomes and only a few with very large incomes. The center figure in Figure 2–6 is negatively skewed. A distribution like this would result if a class were given a very easy test that resulted in a piling up of perfect and near-perfect scores. The bottom figure is symmetrical and is not skewed in either direction. Many psychological and educational variables give such a symmetrical distribution.

Excel and SPSS both provide an index of skewness as part of their descriptive statistics package. The index will yield a positive number if the distribution looks like Figure 2–6a, a negative value if the graph looks like Figure 2–6b, and a value near zero for a distribution that is approximately symmetric.

In the distributions that are approximately symmetrical, either the mean or the median will represent the average of the group equally well, but with skewed distributions, the median generally seems preferable because it is affected less by a few cases out in the long tail. The mean is more often used when the distribution is symmetrical for reasons that will become clear later in this chapter when we discuss the normal distribution. The mode is used less often because, although it is easy to obtain, it is less stable than the mean and median, as we have seen in our example. Also, the mode is not related to other statistics we may wish to obtain.

Figure 2-6

Frequency distributions differing in skewness: (a) positively skewed, (b) negatively skewed, and (c) symmetrical.



MEASURES OF VARIABILITY

It is often significant, when describing a set of scores, to report how *variable* the scores are—that is, how much they spread out from high to low. For example, two groups of children, each with a median age of 10 years, would represent quite different educational situations if one group had a spread of ages from 9 to 11 and the other group ranged from 6 to 14. A measure of this spread is an important statistic for describing a group.

The Range

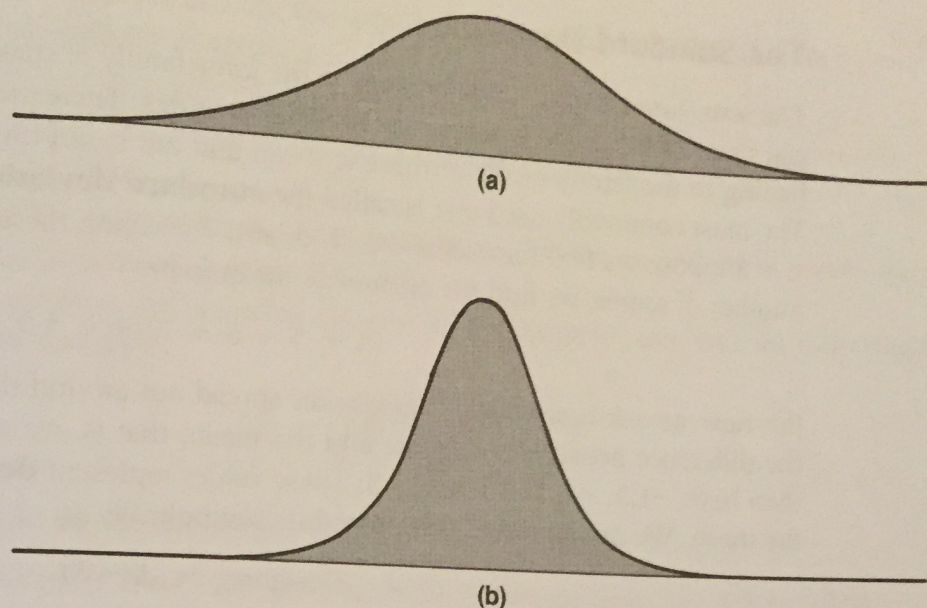
A simple measure of variability is the **range** of scores in the group, which is simply the difference between the highest and the lowest scores. In our math test example, the spread of scores is from 60 to 19, giving a range of 41 points. However, the range depends only on the two extreme cases in the total group. This fact makes the range undependable because it can be changed quite a bit by the addition or omission of a single extreme case. If Uriah Urdahl had not taken the math test, the range would have been 19 to 53, or 34 rather than 41.

The Semi-Interquartile Range

A better measure of variability is the range of scores that includes a specified part of the total group—usually the middle 50%. The middle 50% of the cases in the group are the cases lying between the 25th and the 75th percentiles. We can compute these two percentiles following the procedures outlined earlier. For our example, the 25th percentile was computed to be 32.9. If

Figure 2-7

Two distributions differing only in variability: (a) large variability and (b) small variability.



you calculate the 75th percentile using these procedures, you will find that it is 44.0. The distance between the 25th and 75th percentiles is thus 11.1 points of score, and this range contains the scores of the middle 50% of the distribution, or 26 students.

The 25th and 75th percentiles are called the **quartiles** because they cut off the bottom quarter and the top quarter of the group, respectively. The score distance between them is called the **interquartile range**. A statistic that is often reported as a measure of variability is the **semi-interquartile range (Q)**, which is half of the interquartile range. It is the *average* distance from the median to the two quartiles; that is, it tells how far the quartile points lie from the median on the average. In our example, the semi-interquartile range is

$$Q = \frac{44.0 - 32.9}{2} = 5.55$$

If the middle 26 scores spread out twice as far, Q would be twice as large; if they spread out only half as far, Q would be half as large. Two distributions that have the same mean, the same total number of cases, and the same general form and that differ only in that one has a variability twice as large as the other are shown in Figure 2-7.

MAKING THE COMPUTER DO IT

Quartiles

SPSS will compute the quartiles for you in the Frequencies routine. Click on the box labeled "Quartiles" and be sure you tell the program to treat the data values as group midpoints. For our student math achievement data, we get a very small discrepancy from that obtained above in computing the 25th percentile because SPSS uses a formula that is very general, but the answers will be the same within .01. You must compute the interquartile range and Q for yourself. You can use the Percentile function in Excel to get the 25th and 75th percentiles, but they will be whole points of score because of the way Excel determines percentiles. The quartiles are then used to compute Q .

The Standard Deviation

The semi-interquartile range belongs to the same family of statistics as the median. Both are special cases of the more general concept of percentiles. There are also measures of variability that belong to the family of the arithmetic mean and are based on score deviations from the mean. The most commonly used one is called the **standard deviation**. Let us take a look at it.

Suppose we had four scores: 4, 5, 6, and 7. Adding these scores and dividing by the total number of scores, we find the arithmetic mean to be

$$(4 + 5 + 6 + 7)/4 = 5.5$$

But now we ask how widely these scores spread out around that mean value. Suppose we find the difference between each score and the mean; that is, we subtract 5.5 from each score. We then have -1.5, -0.5, 0.5, and 1.5. These values represent **deviations** of the four scores from the mean. We can represent this procedure symbolically as

$$\text{Deviation} = X - M$$

The bigger the deviations, the more widely the set of scores spreads out around the mean. If our scores were 2, 5, 6, and 9, the mean would still be 5.5, but the deviations would be -3.5, -0.5, 0.5, and 3.5. What we require as an index of the variability, or spread, present in the data is some type of average of these deviations.

If we simply add the four deviation values for either case in the preceding paragraph, we find that they add up to zero. The positive deviations exactly balance the negative ones. This outcome will always be true because one of the definitions of the arithmetic mean is that it is the point around which the sum of deviations is zero. (Note that in our two examples above, the sum of deviations was zero in both cases.) That is

$$\sum(X - M) = 0$$

We will have to do something else to get an index of the amount of spread. The procedure that statisticians have devised for handling the plus (positive) and the minus (negative) signs is to square all the deviations, thus getting only positive values. (A minus times a minus is a plus.) We can obtain an average of these squared deviations by adding them and dividing by the total number of cases. This value is called the **variance**. This statistic is widely used in more advanced statistical procedures. The variance is defined as the *mean* of the *squared* deviations from the mean or

$$\text{Variance} = \frac{\sum(X - M)^2}{N}$$

To compensate for having squared the individual deviations, we must then compute the square root of this average value. The resulting statistic is called the **standard deviation (SD)**. It is the square root of the average of the squared deviations from the mean. (With most calculators, you need only press the designated key to get the square root of a number.) For the first set of scores used in the preceding example, calculations are as follows:

$$\begin{aligned} SD &= \sqrt{\frac{(-1.5)^2 + (-.5)^2 + (.5)^2 + (1.5)^2}{4}} \\ SD &= \sqrt{\frac{2.25 + 0.25 + 0.25 + 2.25}{4}} = \sqrt{\frac{5}{4}} \\ SD &= \sqrt{1.25} = 1.12 \end{aligned}$$

The variance of this set of scores is 1.25; the standard deviation is 1.12. The standard deviation of the second, more variable set of scores is

$$SD = \sqrt{\frac{(-3.5)^2 + (-0.5)^2 + (.5)^2 + (3.5)^2}{4}} = \sqrt{\frac{25}{4}} = 2.5$$

A comparison of the two standard deviations tells us that the second set of scores spreads out more widely around its mean than the first does.

Using the same notation that we encountered with the mean, we can write the following formula for the standard deviation:

$$SD = \sqrt{\frac{\sum(X - M)^2}{N}}$$

This formula tells us to do the following:

1. Subtract the mean from each score to obtain deviations (e.g., $2 - 5.5 = -3.5$).
2. Square each deviation.
3. Sum the squared deviations.
4. Divide the sum of the squared deviations by the number of cases (N).
5. Take the square root of the result.

There is one factor in computing the standard deviation that may cause your calculator or computer to give results that are different from what you would get using the formula given above or that are obtained by a classmate when you are both working with the same set of data. Earlier in the chapter, we mentioned that **statistical inference** was the term applied when using the data from a sample to estimate a characteristic of a larger group (called a **population**). This distinction does not affect the mean (the same value serves as a description of the sample and as an estimate for the population), but it does affect the standard deviation. In the standard deviation that is used to describe the variability of the sample, the sum of the squared deviations from the mean is divided by the number of members of the sample (N). However, the best estimate (from the sample data) of the standard deviation of the population from which the sample comes is found by dividing the sum of squared deviations from the sample mean by $N - 1$ rather than N . Using $N - 1$ in the denominator has the effect of making the population estimate slightly larger than the sample value. The difference is not of practical importance for most measurement applications, but it can cause some confusion when people compare their answers to a problem. You may want to check the manual for your calculator to see which version it uses (many will give you either one) so that you will not get frustrated if you obtain slightly different answers from the ones we provide. For a discussion of why this difference exists and for computing formulas for hand calculators, see any introductory statistics book.

MAKING THE COMPUTER DO IT

Standard Deviation

Most calculators and all computer spreadsheet and statistics packages include a program that will calculate the standard deviation for a set of data. Both Excel and SPSS provide the variance and standard deviation as optional output from several programs. With SPSS, all you have to do is click on the Standard Deviation button in the Statistics section of

Frequencies. If you select the SPSS Descriptives program, the standard deviation is part of the standard output. Excel produces both the standard deviation and variance as part of the output from its Descriptive Statistics routine.

SPSS does not provide an option for which standard deviation you get—it always uses $N - 1$ —but Excel does allow you to get the sample standard deviation if you use the $f_X(\Sigma)$ approach. The function list includes both STDEV (which uses $N - 1$) and STDEVP (which uses N). The difference is minimal unless N is quite small, but knowing that it exists can explain some inconsistencies between summaries of the same data.

INTERPRETING THE STANDARD DEVIATION

It is almost impossible to say in simple terms what the standard deviation is or what it corresponds to in pictorial or geometric terms. Primarily, it is a statistic that characterizes the spread of a distribution of scores. It increases in direct proportion to the scores spreading out more widely around the mean. The larger the standard deviation, the greater the variability among the individuals. A student sometimes asks, but what is a small standard deviation? What is a large one? There is really no answer to either question. Suppose that for some group the standard deviation of weights is 10. Is this value large or small? It depends on whether we are talking about ounces, pounds, or kilograms and on whether we are dealing with the weights of mice, men, or mammoths.

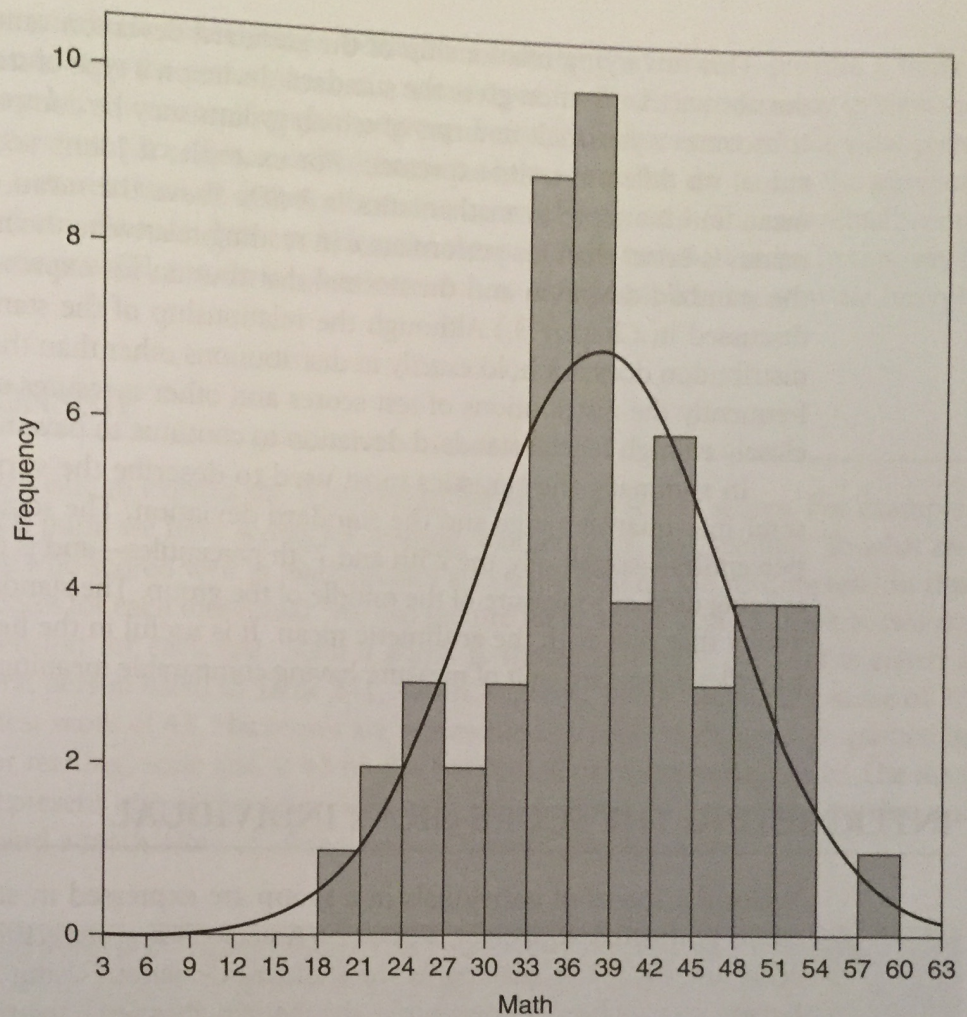
The standard deviation gets its most clear-cut meaning for one type of distribution of scores, the **normal distribution**, or **normal curve**. This distribution is defined by a particular mathematical equation, but to the everyday user, it is defined approximately by its pictorial qualities. The normal curve is a symmetrical curve having a bell-like shape. In fact, it is sometimes called the **bell curve**. Most of the cases pile up in the middle score values; going away from the middle in either direction, the pile drops off, first slowly and then more rapidly and then slowly again as the cases trail out into relatively long tails at each end. An illustration of a typical normal curve is shown in Figure 2-8. This curve is the normal curve that best fits the histogram of the math test data taken from Table 2-1. It has the same mean, standard deviation, and total area (number of cases) as the math test data. The histogram of mathematics test scores (Figure 2-1) is shown as well and reveals how the ideal curve fits the actual test scores. Because the normal curve is symmetrical, the mean, median, and mode all have the same value.

Think back to what the bars of the histogram tell us. Each bar is the same width, and its height is equal to the frequency of scores in the interval covered by the bar. Therefore, the area covered by each bar is proportional to the number of cases in its interval. This means that we can think of the area covered by a part of a graph as equal to the proportion of the scores in our group that fall within the interval. There are $9 + 10 = 19$ individuals falling in the two score intervals 33-35 and 36-38, so there are 19 individuals in this part of the graph. Because there are 52 students in the total group, we can say that $19/52 = .37$ or 37% of them fall in the interval between 33 and 38.

For the normal curve, there is an exact mathematical relationship between the standard deviation and the proportion of cases. The same proportion of cases will always be found within the same standard deviation limits. This relationship is shown in Table 2-5. From this table, we can see that in any normal curve, about two thirds (68.2%) of the cases fall in the range between $+1.0$ SD and -1.0 SD from the mean. Thus, if the mean is 50 and the standard deviation is 10,

Figure 2-8

Normal distribution (bell curve) superimposed on the histogram of 52 mathematics test scores.



about 68% of the cases will fall in the range from a score of 40 to a score of 60 (34% between 40 and 50 and 34% between 50 and 60). Approximately 95% (actually, 95.4%) will fall between $+2.0$ SD and -2.0 SD from the mean, and nearly all the cases will fall between $+3.0$ SD and -3.0 SD from the mean. Because of this constant relationship between the standard deviation and the proportion of cases, we know that in a normal distribution, an individual who gets a score 1 SD above the mean will surpass 84% of the group—the 50% who fall below the mean and the 34% who fall between the mean and $+1.0$ standard deviation.

Table 2-5

Proportion of Cases Falling Within Certain Specified Standard Deviation (SD) Limits for a Normal Distribution

Limits Within Which Cases Lie	% of Cases
Between the mean and <i>either</i> $+1.0$ SD or -1.0 SD	34.1
Between the mean and <i>either</i> $+2.0$ SD or -2.0 SD	47.7
Between the mean and <i>either</i> $+3.0$ SD or -3.0 SD	49.9
Between $+1.0$ SD and -1.0 SD	68.2
Between $+2.0$ SD and -2.0 SD	95.4
Between $+3.0$ SD and -3.0 SD	99.8

This unvarying relationship of the standard deviation unit to the arrangement of scores in the normal distribution gives the standard deviation a type of *standard* meaning as a unit of score. It becomes a yardstick in terms of which groups may be compared or the status of a given individual on different traits expressed. For example, if John's score in reading is 1 SD above the mean and his score in mathematics is 2 SDs above the mean, then his performance in mathematics is better than his performance in reading relative to the mean of each variable. (The use of the standard deviation and the normal distribution for expressing relative performance will be discussed in Chapter 3.) Although the relationship of the standard deviation unit to the score distribution does not hold *exactly* in distributions other than the theoretical normal distribution, frequently the distributions of test scores and other measures approach the normal distribution closely enough for the standard deviation to continue to have nearly the same meaning.

In summary, the statistics most used to describe the variability of a set of scores are the semi-interquartile range and the standard deviation. The semi-interquartile range is based on percentiles—specifically, the 25th and 75th percentiles—and is commonly used when the median is being used as a measure of the middle of the group. The standard deviation is a measure of variability that goes with the arithmetic mean. It is useful in the field of testing primarily because it provides a standard unit of measure having comparable meaning from one test to another.

INTERPRETING THE SCORE OF AN INDIVIDUAL

When the scores of individuals in a group are expressed in standard deviation units, they are called **standard scores** or **Z-scores**. A person's Z-score is the distance between his or her raw score and the mean, divided by the standard deviation. Using X_i to represent the raw score of a person on variable X (for example, the mathematics test), the person's Z-score is

$$Z_{X_i} = \frac{X_i - M_X}{SD_X}$$

The same person's Z-score on the reading test (which we will call variable Y) would be found by

$$Z_Y = \frac{Y_i - M_Y}{SD_Y}$$

If we know that $M_X = 38.17$, $M_Y = 34.44$, $SD_X = 8.93$, and $SD_Y = 5.55$ (these are the means and SDs for the mathematics and reading scores in Table 2-1), then Aaron Andrews' Z-scores on these two variables are

$$Z_X = \frac{43 - 38.17}{8.93} = +0.54$$

$$Z_Y = \frac{32 - 34.44}{5.55} = -0.44$$

(Aaron's raw scores were 43 and 32, respectively.) What these two Z-scores tell us is that Aaron is about one-half standard deviation above the mean in math and about one-half standard deviation below the mean in reading.

The problems of interpreting the score of an individual will be treated more fully in Chapter 3, where we turn to test norms and units of measure. It will suffice now to indicate that the two sorts