Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurements* (4th ed.). Mountain View, CA: Mayfield.

Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist, 30*, 1–14.

DuBois, P. H. (1970). *A history of psychological testing.* Boston: Allyn & Bacon.

Gottfredson, L. S., & Sharf, J. C. (1988). *Fairness in employment testing: A special issue of the Journal of Vocational Behavior, 33*(3). Duluth, MN: Academic Press.

Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn & Bacon.

Haney, W. (1981). Validity, vaudeville, and values: A short history of social concerns over standardized testing. *American Psychologist, 36*, 1021–1034.

Hartigan, J. A., & Wigdor, A. K. (Eds.) (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Howard, G. S. (1985). The role of values in the science of psychology. *American Psychologist, 40*, 255–265.

Jones, L. V. (1971). The nature of measurement. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 335–355). Washington, DC: American Council on Education.

Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 471–516). Westport, CT: Praeger.

Linn, R. L. (1989). Current perspectives and future directions. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 1–12). New York: Macmillan.

Murphy, K. R., & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Rogers, T. B. (1995). *The psychological testing enterprise.* Pacific Grove, CA: Brooks/Cole.

Thorndike, R. M. (1990). *A century of ability testing.* Chicago: Riverside.

Vold, D. J. (1985). The roots of teacher testing in America. *Educational Measurement: Issues and Practice, 4*(3), 5–8.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982). *Ability testing: Uses, consequences, and controversies: Pt. I. Report of the committee.* Washington, DC: National Academy Press.

CHAPTER 2

# Measurement and Numbers



## QUESTIONS TO ASK ABOUT TEST SCORES

Catherine Johnson and Peter Cordero wanted to gather information about achievement levels in their two sixth-grade classes. They gave their students a 45-item reading comprehension test provided in their current reading series, a 65-item review test from the mathematics book, and a dictation spelling test of 80 items based on the words their classes had been studying during the past 6 weeks. They marked the papers, counted the number of correct answers on each, and recorded the scores. Then they made up a joint class list that showed the three scores for each student. In addition, they recorded which class each student was in (1 for Ms. Johnson's class, 2 for Mr. Cordero's) and each student's gender (1 for boys, 2 for girls). With five pieces of information for each of 52 students, they wondered what they should do with all those numbers.

Tests and assessments *do* produce scores, and scores *are* numbers. So, if we are to think about and use test scores, we must be prepared to think about and work with numbers. The numbers that represent test scores can be organized to provide the answers to a range of questions, but first we must know what kinds of questions to ask. Once we have the questions in mind, we can begin to ask how the numbers can be arranged to provide the answers.

Look at the numbers (scores) shown in Table 2–1 for the sixth-graders in the two classes. What kinds of questions might the two teachers ask about this set of numbers?

**Table 2–1**
Scores for 52 Sixth-Grade Students on Tests of Reading, Spelling, and Mathematics

| First Name | Last Name | Gender | Class | Reading (45)* | Spelling (80) | Math (65) |
|---|---|---|---|---|---|---|
| Aaron | Andrews | 1 | 1 | 32 | 64 | 43 |
| Byron | Biggs | 1 | 1 | 40 | 64 | 37 |
| Charles | Cowen | 1 | 1 | 36 | 60 | 38 |
| Donna | Davis | 2 | 1 | 41 | 74 | 40 |
| Erin | Edwards | 2 | 1 | 36 | 69 | 28 |
| Fernando | Franco | 1 | 1 | 41 | 67 | 42 |
| Gail | Galaraga | 2 | 1 | 40 | 71 | 37 |
| Harpo | Henry | 1 | 1 | 30 | 51 | 34 |
| Irrida | Ignacio | 2 | 1 | 37 | 68 | 35 |
| Jack | Johanson | 1 | 1 | 26 | 56 | 26 |
| Kleven | Klipsch | 1 | 1 | 28 | 51 | 25 |
| Laverne | Lappenski | 2 | 1 | 36 | 57 | 53 |
| Mary | Madison | 2 | 1 | 39 | 68 | 37 |
| Nathan | Natts | 1 | 1 | 22 | 47 | 22 |
| Oprah | Oates | 2 | 1 | 36 | 59 | 33 |
| Petula | Peters | 2 | 1 | 32 | 64 | 33 |
| Quadra | Quickly | 2 | 1 | 21 | 44 | 19 |
| Rahim | Roberts | 1 | 1 | 29 | 64 | 43 |
| Salim | Salik | 1 | 1 | 41 | 76 | 33 |
| Thomas | Tank | 1 | 1 | 35 | 65 | 38 |
| Usaka | Urban | 2 | 1 | 41 | 65 | 38 |
| Victor | Vasquez | 1 | 1 | 37 | 68 | 40 |
| Wakana | Watanabe | 2 | 1 | 25 | 53 | 21 |
| Xenum | Xerxes | 1 | 1 | 25 | 54 | 31 |
| Yuan | Young | 1 | 1 | 32 | 59 | 24 |
| Zebulon | Zibberits | 1 | 1 | 42 | 73 | 44 |
| Angela | Ash | 2 | 2 | 43 | 64 | 52 |
| Bellinda | Brown | 2 | 2 | 33 | 38 | 41 |
| Charlotta | Cowen | 2 | 2 | 33 | 47 | 50 |
| Dominik | Dubrow | 1 | 2 | 39 | 66 | 34 |
| Erik | Eriksen | 1 | 2 | 39 | 55 | 47 |
| Francis | French | 2 | 2 | 38 | 59 | 49 |
| Guido | Garcia | 1 | 2 | 31 | 52 | 29 |
| Hiliary | Huan | 2 | 2 | 38 | 61 | 48 |
| Igor | Ivanovich | 1 | 2 | 33 | 53 | 43 |

*(continued)*



**Table 2–1 *(Continued)***

| First Name | Last Name | Gender | Class | Reading (45)* | Spelling (80) | Math (65) |
|---|---|---|---|---|---|---|
| Jill | Johanson | 2 | 2 | 42 | 61 | 45 |
| Kaleen | Knowles | 2 | 2 | 35 | 55 | 51 |
| Larry | Lewis | 1 | 2 | 29 | 40 | 34 |
| Moe | Mastrioni | 1 | 2 | 36 | 58 | 39 |
| Nancy | Nowits | 2 | 2 | 28 | 44 | 44 |
| Orden | Orford | 1 | 2 | 35 | 53 | 38 |
| Petre | Popovich | 1 | 2 | 36 | 52 | 53 |
| Quincy | Quirn | 1 | 2 | 33 | 48 | 33 |
| Rhonda | Rostropovich | 2 | 2 | 31 | 50 | 31 |
| Sally | Stebbens | 2 | 2 | 33 | 51 | 32 |
| Thelma | Thwaites | 2 | 2 | 38 | 43 | 45 |
| Uriah | Urdahl | 1 | 2 | 42 | 61 | 60 |
| Velma | Vauter | 2 | 2 | 29 | 49 | 36 |
| William | Westerbeke | 1 | 2 | 33 | 54 | 33 |
| Xena | Xerxes | 2 | 2 | 30 | 57 | 37 |
| Yannita | Younts | 2 | 2 | 44 | 63 | 49 |
| Zephina | Zoro | 2 | 2 | 30 | 47 | 38 |

*Maximum score in parentheses.

What questions can *you* ask? Before reading further, study the sets of scores and jot down the questions that come to your mind in connection with these scores. See how many of the question types that we discuss you can anticipate.

Each student has five numbers assigned to him or her, but we might want to inquire whether each of these numbers actually conveys quantitative information. Do the numbers that are assigned to represent gender and class have the same kind of meaning as the test scores? This is a question of the *scale* the numbers represent, and the answer affects the kinds of operations we can apply to the numbers.

A second, rather general type of question that we might ask is, what is the basic pattern of the set of scores? How do they "run"? What do they "look like"? How can we get a picture of the set of math scores, for example, so that we can get an impression of the group as a whole? To answer this type of question, we will need to consider simple ways of tabulating and graphing a set of scores.

A third type of question that will almost certainly arise is, what is this group like, on the average? In general, have they done as well on the test as some other sixth-grade group? What is the typical level of performance in the group? All these questions call for some single number to represent the group as a whole, some measure of where the *middle* of the group lies. To answer this type of question, we will need to become acquainted with statistics developed to represent the average, or typical, score.

Fourth, to describe the group, we might feel a need to describe the extent to which the scores spread out away from the average value. Have all the students in the group made about the same progress, or do they show a wide range of achievement? How does this group compare with other classes, with respect to the *spread* of scores, and do the students show the same spread of achievement on all three tests? This type of question calls for a study of measures of variability.

Fifth, we might ask how a particular individual stands on one of the tests. We might want to know whether Aaron Andrews did well or poorly on the mathematics test. And if we decide that his score is a good one, we might want some way of saying just how good it is. We might ask whether Aaron did better in reading or in mathematics. To answer this question, we will need a *common yardstick* on which to express performance in two quite different areas. One need, then, is for some uniform way of expressing and interpreting the performance of an individual, independent of the particular test. How does this person stand relative to the group? We will give some preliminary answers to this question in this chapter and consider it in detail in Chapter 3.

A sixth query is of the following type: To what extent do those who excel in reading also excel in mathematics? To what extent do these two abilities go together in the same individuals? Is the individual who is superior in one area likely to be superior in the other? To express this *association* between two measurements, we will need to become acquainted with indices of correlation.

Test scores are also frequently used to forecast future performance. Therefore, a seventh issue we might want to address is how to make the most accurate prediction of a person's performance either on another test or on some outcome, such as an end-of-the-year achievement assessment. This forecasting function is the primary use that is made of scores on such high-stakes tests as the SAT, and the answer will lead us to a consideration of regression.

Many other questions may arise with respect to a set of scores. The most important ones concern the drawing of general conclusions from data on a limited group. For example, the 26 girls in this group have an average reading score of 35.0, and the 26 boys have an average score of 33.9. These are *descriptive* facts about this testing of these particular girls and boys. But these students might be considered to represent a larger population, such as all students in this school district or the state. From the results in these two classes, we might want to make an estimate or best guess of the average level of reading achievement in the larger group. We also might want to know whether we can safely conclude that the total population of girls from which this sample is drawn would surpass the total population of boys on this same test. These problems are of **inference.** Problems of statistical inference make up the bulk of advanced statistical work. For a detailed description of the principles and applications of statistical inference, see any of the books on statistical methods listed at the end of this chapter. These issues do not enter into the basic interpretation of test scores for an individual or a group, so we will not consider them further here.

The routines developed for organizing numbers to answer these and other questions constitute the field called **statistics.** This name and, in fact, the very prospect of working with numbers seem a bit scary to some people. Fortunately, much of the mechanics of working with numbers can now be performed on a pocket calculator or personal computer, so we can concentrate on the questions to ask and on the ways in which the numbers are arranged to answer them, rather than worry about computational details. As we discuss ways to answer each of the seven types of questions mentioned above, we will introduce easy-to-use computer programs to perform the necessary computations. Two very widely available programs will be presented: the package of statistical programs known as SPSS and the data analysis routines included in the Microsoft Excel® spreadsheet program. Both programs have some shortcomings, but both are relatively easy to use once you get used to them.



## SCALES OF MEASUREMENT

One way to define measurement is *the assignment of numbers to objects according to a set of rules.* The set of rules is called a **scale.** Knowing the scale that has been used to assign the numbers is critical to proper interpretation of the measurement. For example, Ms. Johnson and Mr. Cordero assigned the number 1 to their male students and the number 2 to their female students, but what information do these numbers contain? Feminist ideology aside, does the number 2 mean that the girls possess more of the trait of gender than the boys do? Obviously not; in this case the numbers do not convey information about *amount* of anything. The number 1 has been substituted for the label "boy" and the number 2 has been substituted for "girl," but we could just as well have used the numbers 163 and 27. Each number takes on the meaning of a verbal label. When numbers are used in this way, the scale is called a **nominal scale.** The numbers take the place of names. Your university has almost certainly assigned you a student number. This number represents a "score" on a nominal scale, as do the numbers on the backs of athletes' uniforms. When numbers do not contain information about amount of a trait, it is not appropriate to treat them like numbers. In general, you cannot add them or perform any other arithmetic operations and obtain a meaningful result. The only use we ordinarily can make of numbers on a nominal scale is to count how many instances there are of each number. There are 26 ones and 26 twos in Table 2–1, so we know there are an equal number of boys and girls.

Sometimes numbers are assigned to represent the order of individuals on a trait. None of the numbers in Table 2–1 are of this kind, but suppose Mr. Cordero decided to rank order his students on the basis of their spelling test scores with the student who earned the highest score—Dominik Dubrow, getting a rank of 1, second highest, Angela Ash, ranked 2, and so forth. What kind of information would these numbers represent?

A set of ranks conveys information about the order in which the students stand on the trait, but the ranks do not contain information about amount. More importantly, the differences between numbers do not have a constant meaning. The difference between a rank of 1 and a rank of 5 is not necessarily the same as the difference between a rank of 11 and one of 15. Both differ by four units, but four units of rank usually cover a greater portion of the trait at the extremes of a group than in the middle. A scale that tells us the order in which people stand, who has more of the trait, but not how much more, is called an **ordinal scale.** Several common ways of reporting test score information that we will describe in Chapter 3 produce ordinal scales.

Tests that are scored like those in Table 2–1 treat each item as equal to every other item. Getting 10 items correct yields a score of 10, getting 20 items correct yields a score of 20, and getting 30 items correct would give you a score of 30. Because each item is assumed equal in amount of the trait it measures, equal differences in scores are treated as equal differences in the trait. Although this assumption is somewhat tenuous when measuring human abilities, it represents the same basic kind of measurement as the centigrade or Fahrenheit temperature scales. Equal numerical differences in score represent equal differences in the property being measured. When we can make this assumption, the scale is called an **interval scale.** Most of the computations that are done with test scores require that we assume the scale of measurement is an interval scale.

Suppose Zebulon Zibberits tries his hardest, but still cannot spell any of the words in Mr. Cordero's list correctly. Does this mean that Zebulon has zero spelling ability? Probably not. The scale of spelling ability represented by this test starts at a point well above zero, so a score of zero on the test does not mean zero ability. Very few scales used in psychology and education are constructed such that a score of zero means exactly none of the trait or property in question. This is one reason Blanton and Jaccard (2006) expressed concern about the arbitrary nature of the metrics used in psychology and education. Exceptions (non-arbitrary metrics) would be scales like those for height, weight, and duration of time. When the scale is constructed so a score of zero means exactly none of the trait, it is called a **ratio scale.** Scales like this allow us to conclude not only that the size of the unit is the same everywhere along the scale, but also that a score that is numerically twice another score means exactly twice as much of the trait. Someone who takes 10 minutes to solve a problem takes twice as long as someone who takes 5 minutes, and the person who takes 20 minutes to solve the problem takes twice as long as the 10-minute person and four times as long as the 5-minute person. Ratio scales allow us to make proportional statements like these. Very few ratio scales exist in the fields of education and psychology, but fortunately interval scales allow us to perform most of the analyses we need. This classification of scales as nominal, ordinal, interval or ratio was suggested by S. S. Stevens in 1951 and has been widely accepted, although some experts in measurement, for example, Torgerson (1958) have suggested alternative classifications.

## PREPARATION OF A FREQUENCY DISTRIBUTION

In Table 2–1, we showed a record sheet on which test scores for 52 sixth-graders were listed. Let us look at the scores in the Math column and consider how they can be rearranged to give a clearer picture of how the pupils have performed on the math test. The simplest rearrangement is merely to list the scores in order from highest to lowest, as follows:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 60 | 49 | 44 | 41 | 38 | 37 | 33 | 31 | 24 |
| 53 | 49 | 44 | 40 | 38 | 36 | 33 | 31 | 22 |
| 53 | 48 | 43 | 40 | 38 | 35 | 33 | 29 | 21 |
| 52 | 47 | 43 | 39 | 37 | 34 | 33 | 28 | 19 |
| 51 | 45 | 43 | 38 | 37 | 34 | 33 | 26 | |
| 50 | 45 | 42 | 38 | 37 | 34 | 32 | 25 | |

This arrangement gives a somewhat better picture of the way the scores fall than does Table 2–1. We can see the highest (60) and lowest (19) scores at a glance. It is also easy to see that the middle person in the group falls somewhere in the mid-30s. We can see by inspection that most of the scores fall between 30 and 50. But this simple rearrangement of scores still has too much detail for us to see the general pattern clearly. We need to condense the data into a more compact form.

Often, the first step in organizing scores for presentation is to prepare a display called a **frequency distribution,** a table that shows how often each score has occurred. Each score value is listed, and the number of times it occurs is shown. A portion of the frequency distribution for the math test is shown in Table 2–2. However, Table 2–2 is still not a very good form for reporting the scores. The table is too long and spread out. We have shown only part of it; the whole table would take 42 lines, almost as many as the original listing of scores. It would have a number of zero entries, and there would be marked variations in the Frequency column from one score to the next.



**Table 2–2**
Frequency Distribution of Scores on the Mathematics Test for 52 Students

| Score (*X*) | Frequency |
|---|---|
| 60 | 1 |
| 59 | 0 |
| 58 | 0 |
| 57 | 0 |
| 56 | 0 |
| 55 | 0 |
| 54 | 0 |
| 53 | 1 |
| 52 | 1 |
| . | . |
| . | . |
| . | . |
| 40 | 2 |
| 39 | 1 |
| 38 | 5 |
| 37 | 4 |
| . | . |
| . | . |
| . | . |
| 28 | 1 |
| 27 | 0 |
| 26 | 1 |
| 25 | 1 |
| 24 | 1 |
| 23 | 0 |
| 22 | 1 |
| 21 | 1 |
| 20 | 0 |
| 19 | 1 |

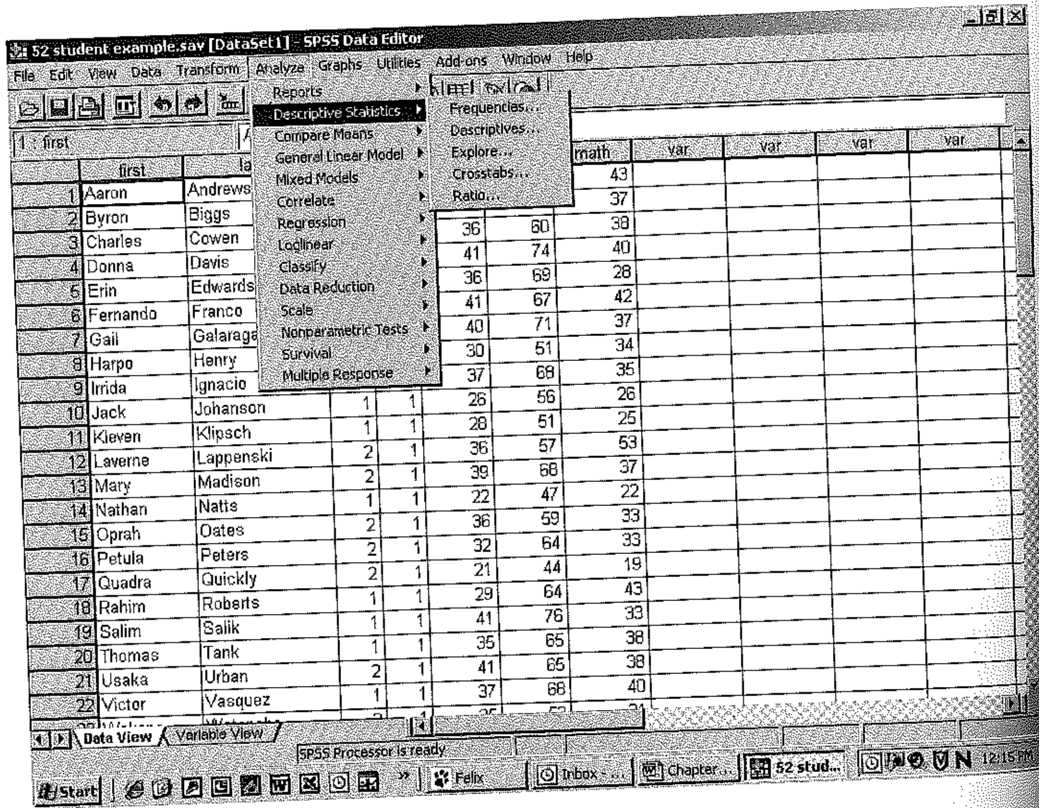
### Grouped Frequency Distributions

Scores are often *grouped* into broader categories to further improve the clarity of presentation. We discard some detail in the data to make it easier to grasp the picture presented by the entire set of scores. In our example, we will group three adjacent scores, so that each grouping *interval* includes three points of score. The entire range of scores from 19 to 60 is

## MAKING THE COMPUTER DO IT

### Frequency Distributions

Excel is not designed to provide a frequency distribution like the one we have just described. Rather, it will prepare a grouped frequency distribution as described in the next section. SPSS will produce a frequency distribution, but you must be careful when reading it because score values with zero frequency are omitted from the list of scores. To prepare a frequency distribution of the data from Table 2–1 using SPSS, start the program, enter the data from Table 2–1 or open the file in which they have been saved, click on the Analyze button in the menu at the top of the screen, and select Descriptive Statistics. The screen should look like the one shown below (the exact layout of the screen will depend on which version of SPSS you are using):




Select Frequencies and you will see a dialogue box in which you can select the variables to be analyzed. Highlight the name of the variable you want by clicking on it, then click on the arrow button to move the variable into the Variables window. If you use the data from Table 2–1 and select the Math variable, you should get the complete frequency distribution, part of which we showed you in Table 2–2. You can generate other reports from this program, but we will wait to describe them until we have covered those topics.



**Table 2–3**
Grouped Frequency Distribution of Scores from 52 Students on a Math Test Using an Interval of 3

| Interval | Frequency |
|---|---|
| 60–62 | 1 |
| 57–59 | 0 |
| 54–56 | 0 |
| 51–53 | 4 |
| 48–50 | 4 |
| 45–47 | 3 |
| 42–44 | 6 |
| 39–41 | 4 |
| 36–38 | 10 |
| 33–35 | 9 |
| 30–32 | 3 |
| 27–29 | 2 |
| 24–26 | 3 |
| 21–23 | 2 |
| 18–20 | 1 |

represented by 14 intervals. When this is done, the set of scores is represented as shown in Table 2–3, a fairly compact table illustrating how many people there are in each **score interval.** Thus, for example, we have two people in the interval 19–21. We do not know how many of them got 19s, 20s, or 21s; we have lost this information in the grouping. *We assume that they are evenly spread throughout the interval.* In most cases, there is no reason to believe that one score will occur more often than any other, and this assumption is a sound one, so the gains in compactness and convenience of presentation more than make up for any slight inaccuracy introduced by the groupings. (In some special applications, such as reports of family income, certain values are more likely than others, for example, $18,000, $25,000, $50,000. Special precautions are required when grouping material of this type. An effort should be made to place the most popular values near the middle of an interval to reduce distortion.)

In practical situations, we always face the problem of deciding how broad the groupings should be, that is, whether to group by 3, 5, 10, or some other number of points of score. The decision is a compromise between (1) losing detail from our data and (2) obtaining a convenient, compact, and smooth representation of the results. The use of broader intervals results in losing more detail, but condenses the data into a more compact picture. A practical rule of thumb is to choose an interval that will divide the total score range into roughly 15 groups. In our example, the highest score is 60, and the lowest is 19. The range of scores is 60 to 19, giving a range of 41 points. Dividing 41 by 15, we get 2.7. The nearest whole number is 3, so we group the data by 3s. In addition to the "rule of 15," we also find that intervals of 5, 10, and multiples of 10 make convenient groupings. Because the purpose of grouping scores is to arrive at a convenient and clear representation, factors of convenience become a major consideration. It is also conventional to use a multiple of the interval width as the lower limit for each interval; as we have done in Table 2–3.

In cases where graphs are going to be prepared using the scores in their grouped form, it is also convenient to use an interval that includes an odd number of score points, for example, 3, 5, or 7, because it is sometimes necessary to use the midpoint of the score interval to represent all scores in the interval. If the interval has an even number of score values, this midpoint will be halfway between two actual scores, but if an odd number is used for the interval width, the midpoint will be a whole score value, making for a more attractive graph.

Note also that sometimes there is no need to group the data into broader categories. If the original scores cover a range of no more than about 20 points, grouping may not be required. Also, with modern computing equipment it is usually easier to compute the statistical indices, described later in the chapter, from the original set of scores unless the data come to you as a grouped frequency distribution. In that case, most programs require special procedures that are beyond our scope.

## MAKING THE COMPUTER DO IT

### Grouped Frequency Distributions

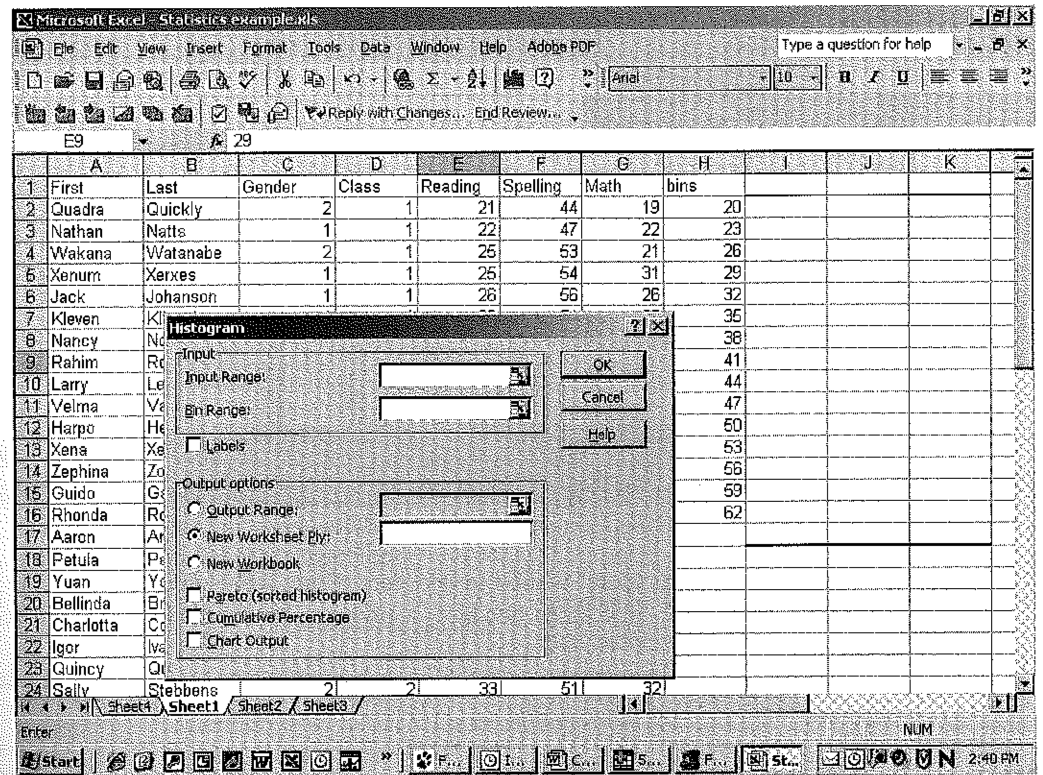
Excel is designed to produce grouped frequency distributions, but you must follow a particular sequence of steps that can at times be frustrating until you become used to them. After you have opened the file containing the data you wish to analyze, the first step is to decide on the number of intervals. Excel calls these intervals *bins*. Using our set of scores for the math test, the easiest way to determine how many bins to use is to sort the scores to find the range they cover. Click on any score in the Math column, then click on one of the sort icons. The scores will now be in order, so you can determine the highest and lowest. Finding the range to be 41 and applying our "rule of 15," we again decide to use bins of 3. (Note that you could use Excel to prepare a raw frequency distribution like the one we got using SPSS, but you would have to create a bin for every possible score.)

You must now set up your bins. Bins are specified by giving the *highest* score that is to fall in each bin. Therefore, if we want to use a bin size of 3, our lowest bin should have 20 specified as its value (to make the lower limit of each bin a multiple of 3). Excel sorts scores into bins in ascending order. That is, the program looks at all the scores and puts those that are less than or equal to the bin limit in the first bin. If we start with a bin limit that is below the lowest score, it will have a frequency of zero, but if we start with a bin limit that is more than our chosen interval above the lowest score, the lowest interval will be too wide.

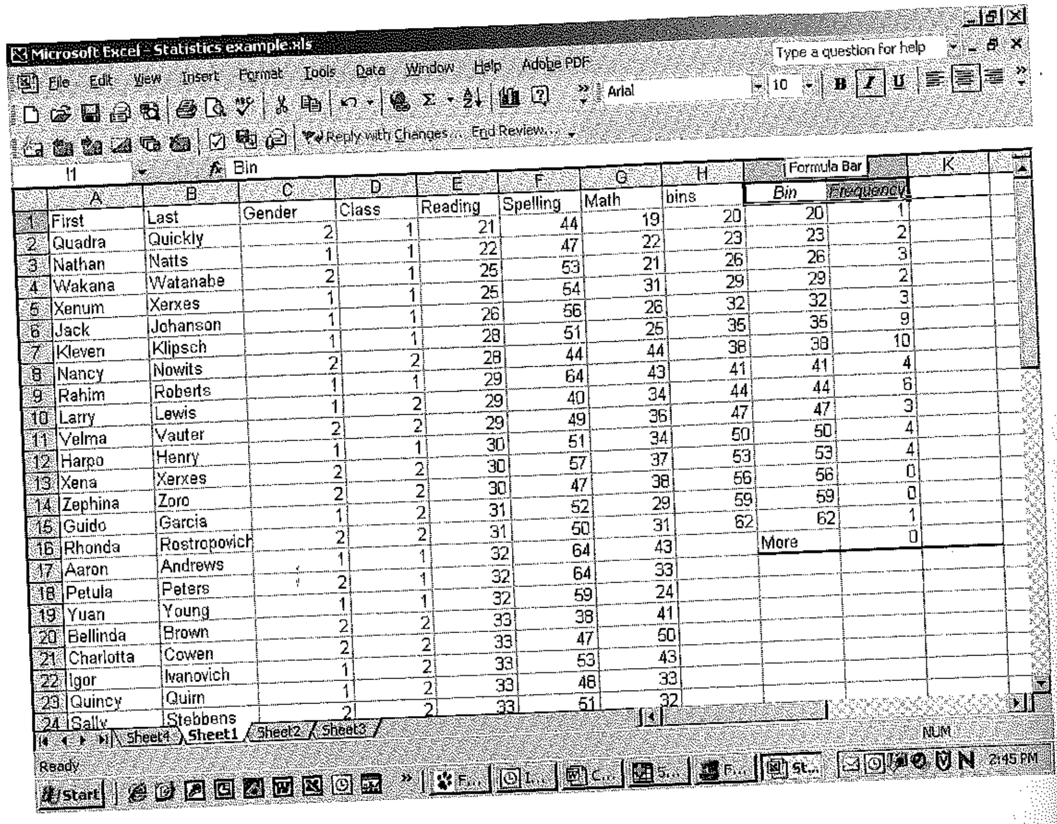
After deciding on the first bin limit, set each additional bin limit one interval above the one below it. If we start with a limit of 20, then the other bin limits would be 23, 26, 29, . . . 53, 56, 59, 60. *The bins must be arranged with the lowest bin at the top of its column*. In a column to the right of your data (for instance, column I), type the word *bins* in the first row. Then type in the bin limits, starting with the lowest, in the cells under the "bins" heading.



The Excel frequency distribution routine is part of the Histogram program in the Data Analysis group of programs that are included in the Tools menu. (Histograms are discussed in the next section.) Click on Tools, then on Data Analysis. (*Hint:* You may have to click on a cell containing data to activate the Data Analysis tool.) A dialogue box will appear that includes many data analysis options. Click on Histogram and click OK. You should see a screen that looks like the one shown here imposed over your data.




Excel works by specifying areas of the data table to be included in the analysis. The scores for the math test are listed in column G with a label in the first row and data in cells G2 to G53. Click on the Input Range window in the Histogram pop-up dialogue box. Then in the main screen highlight the cells containing the math test scores. The cell references $G$2 and $G$53 will appear in the pop-up box window. Next, click in the Bin Range window of the pop-up box. Then, again on the main screen, highlight the bin limits you placed in the column under "bins." The cell references for these cells will appear in the window. Finally, click the Output Range button on the pop-up box to activate that window and *click in the window*. Then click in a cell where you would like the grouped frequency distribution to start (usually just to the right of your bins column, in this case $I$1). Then click OK. If you are using the data from Table 2–1 and have followed the directions as we have outlined them, your screen should look like this:




| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | First | Last | Gender | Class | Reading | Spelling | Math | bins | *Bin* | *Frequency* |
| 2 | Quadra | Quickly | 2 | 1 | 21 | 44 | 19 | 20 | 20 | 1 |
| 3 | Nathan | Natts | 1 | 1 | 22 | 47 | 22 | 23 | 23 | 2 |
| 4 | Wakana | Watanabe | 2 | 1 | 25 | 53 | 21 | 26 | 26 | 3 |
| 5 | Xenum | Xerxes | 1 | 1 | 25 | 54 | 31 | 29 | 29 | 2 |
| 6 | Jack | Johanson | 1 | 1 | 26 | 56 | 26 | 32 | 32 | 3 |
| 7 | Kleven | Klipsch | 1 | 1 | 28 | 51 | 25 | 35 | 35 | 9 |
| 8 | Nancy | Nowits | 2 | 2 | 28 | 44 | 44 | 38 | 38 | 10 |
| 9 | Rahim | Roberts | 1 | 1 | 29 | 64 | 43 | 41 | 41 | 4 |
| 10 | Larry | Lewis | 1 | 2 | 29 | 40 | 34 | 44 | 44 | 6 |
| 11 | Velma | Vauter | 2 | 2 | 29 | 49 | 36 | 47 | 47 | 3 |
| 12 | Harpo | Henry | 1 | 1 | 30 | 51 | 34 | 50 | 50 | 4 |
| 13 | Xena | Xerxes | 2 | 2 | 30 | 57 | 37 | 53 | 53 | 4 |
| 14 | Zephina | Zoro | 2 | 2 | 30 | 47 | 38 | 56 | 56 | 0 |
| 15 | Guido | Garcia | 1 | 2 | 31 | 52 | 29 | 59 | 59 | 0 |
| 16 | Rhonda | Rostropovich | 2 | 2 | 31 | 50 | 31 | 62 | 62 | 1 |
| 17 | Aaron | Andrews | 1 | 1 | 32 | 64 | 43 | | More | 0 |
| 18 | Petula | Peters | 2 | 1 | 32 | 64 | 33 | | | |
| 19 | Yuan | Young | 1 | 1 | 32 | 59 | 24 | | | |
| 20 | Bellinda | Brown | 2 | 2 | 33 | 38 | 41 | | | |
| 21 | Charlotta | Cowen | 2 | 2 | 33 | 47 | 50 | | | |
| 22 | Igor | Ivanovich | 1 | 2 | 33 | 53 | 43 | | | |
| 23 | Quincy | Quirn | 1 | 2 | 33 | 48 | 33 | | | |
| 24 | Sally | Stebbens | 2 | 2 | 33 | 51 | 32 | | | |

The row labeled "More" is used to take care of any scores that are above the highest bin limit you listed. SPSS does not have a program for producing grouped frequency distributions.

## Cumulative Frequency Distributions

We often want to know how many people got scores below some particular value. The most direct way to answer this question is with a **cumulative frequency distribution,** which lists each score or interval and the number of scores falling *in or below* the score or interval. A cumulative frequency distribution is easily prepared from the frequency distribution or grouped frequency distribution, as shown in Table 2–4, which presents the **cumulative frequency,** as well as the frequency in each interval. Each entry in the column labeled "Cumulative Frequency" shows the total number of individuals having a score **equal to or less than** the highest score in that interval; that is, there is 1 student scoring *at or below* 20, (1 + 2) = 3 students scoring at or below 23, (3 + 3) = 6 scoring at or below 26, (6 + 2) = 8 scoring at or below 39, (8 + 3) = 11 scoring at or below 32, and so forth. The cumulative frequency distribution is especially useful for determining some expressions of relative position, which we will discuss in Chapter 3. Unfortunately, neither SPSS nor Excel has a routine for providing cumulative frequency distributions,



**Table 2–4**
Cumulative Frequency Distribution of Scores from 52 Students on a Math Test Using an Interval of 3

| Interval | Frequency | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|
| 60–62 | 1 | 52 | 100 |
| 57–59 | 0 | 51 | 98 |
| 54–56 | 0 | 51 | 98 |
| 51–53 | 4 | 51 | 98 |
| 48–50 | 4 | 47 | 90 |
| 45–47 | 3 | 43 | 83 |
| 42–44 | 6 | 40 | 77 |
| 39–41 | 4 | 34 | 65 |
| 36–38 | 10 | 30 | 58 |
| 33–35 | 9 | 20 | 38 |
| 30–32 | 3 | 11 | 21 |
| 27–29 | 2 | 8 | 15 |
| 24–26 | 3 | 6 | 12 |
| 21–23 | 2 | 3 | 6 |
| 18–20 | 1 | 1 | 2 |

but both will give you cumulative percents; that is, each frequency distribution program has the ability to output the values equal to

$$\text{Cumulative percent} = \frac{\text{cumulative frequency}}{\text{total number of cases}}$$

These values are included in Table 2–4. The cumulative percents are given whenever you request a frequency distribution from SPSS. With Excel you must click on the Cumulative Percentages box in the Histogram dialogue box.

# GRAPHIC REPRESENTATION

It is often helpful to translate the facts of a table like Table 2–3 into a pictorial representation. A common type of graphic representation, called a **histogram,** is shown in Figure 2–1. This type of graph can be thought of, somewhat grimly, as "piling up the bodies." The score intervals for the mathematics test scores that we used in Table 2–3 are shown along the baseline (the **abscissa**), and the vertical height of the pile (the **ordinate**) represents the number of people whose scores fall in that interval. The diagram shows that there is one "body" piled up in interval 18–20, two in interval 21–23, and so forth. (Because we used Excel to generate this histogram, the intervals are labeled with the upper bin limits.) This figure gives a clear picture of how the scores pile up, with most of them in the 30 to 45 range and long, low "tails" running out to the extreme low and high scores.

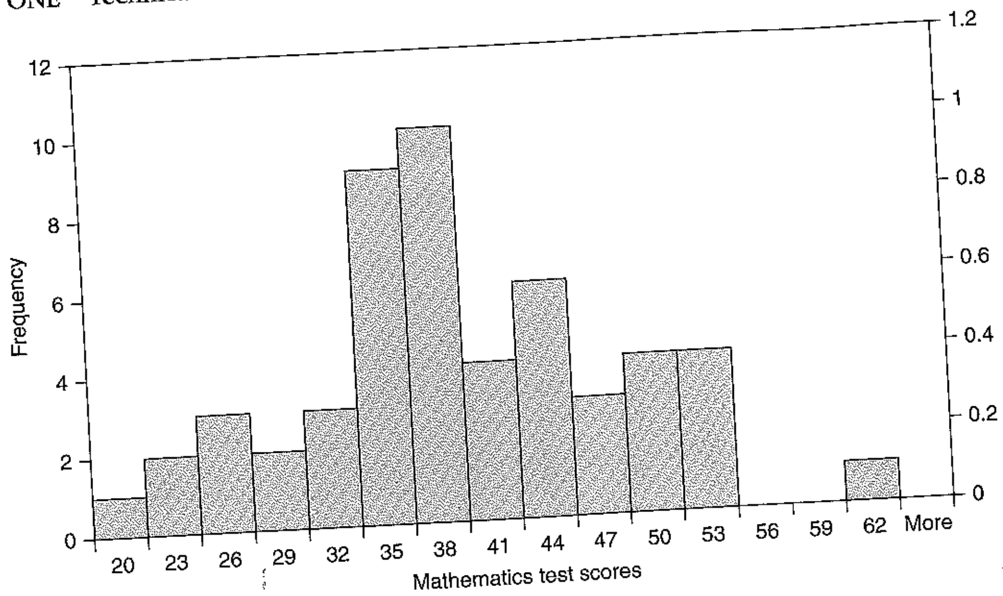



**Figure 2–1**
Histogram of 52 mathematics scores.

## MAKING THE COMPUTER DO IT

### Histograms

Both SPSS and Excel will prepare graphs of grouped frequency distributions, but they do it in very different ways. Excel's procedure is by far the simpler one, but from the technical point of view, the graph Excel's default options produce is properly called a **bar graph**, not a histogram because the bars are separated from each other. This can be corrected by double clicking on any bar, selecting "Options" and changing the "gap width" to zero. Unfortunately, the intervals are labeled with the upper limit of each "bin" rather than the interval midpoint, which may cause problems in graph interpretation if this is not pointed out. The lesson in this example is that you have to look at computer output with caution because the choices made by the programmer who wrote the program may not give you the exact picture you expect or want.

SPSS automatically groups the data to produce a graph with about 15 categories (the default for our example yields 13). The result is a proper histogram (the bars touch each other), but you may want to edit the graph to improve its appearance. If you double-click on the output of the graph, you can edit it. You may want to experiment with the graph editor to see what the editing features produce. The SPSS graph, on page 37, of the mathematics data has been edited to make it as similar as possible, within SPSS editing options, to the graph of our grouped frequency distribution. However, this graph does



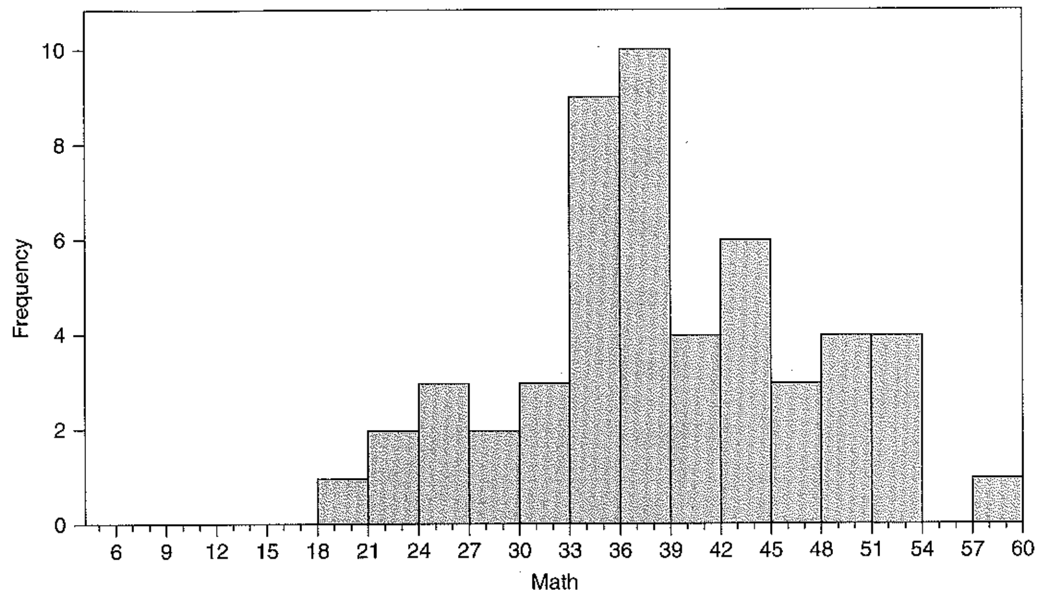


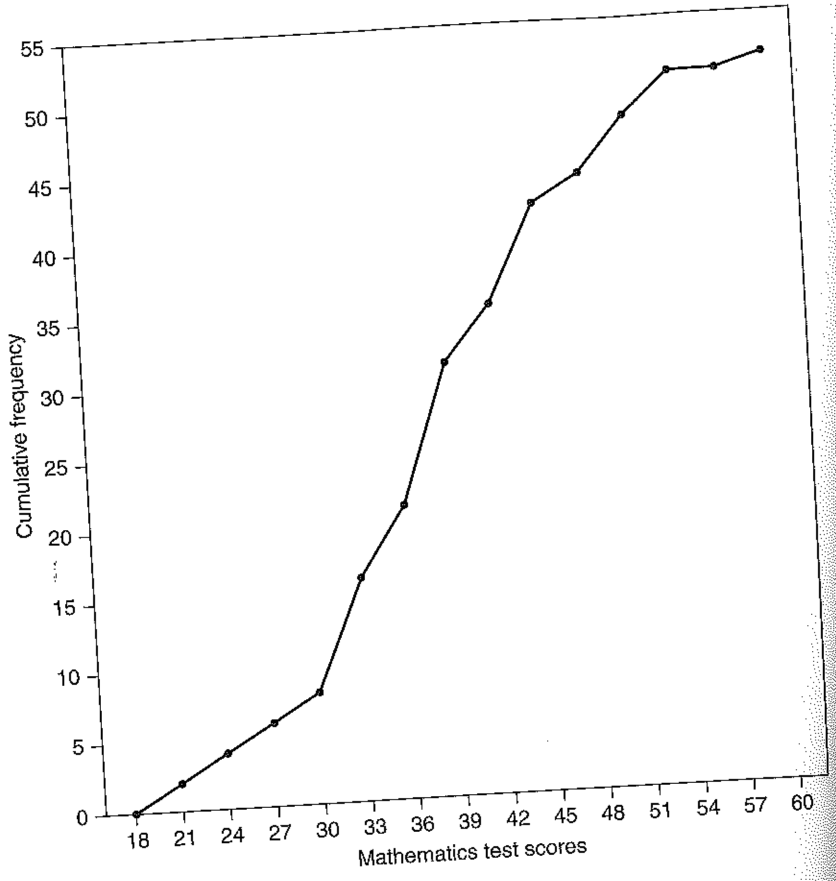

**Figure 2–2**
Histogram of math scores produced by SPSS.

not represent the grouped frequency distribution shown in Table 2–4 and pictured in Figure 2–2.

One peculiarity of SPSS is that it treats the scale of the data as strictly continuous, which can create problems. Default interval widths and limits often are not obtainable score values, resulting in interval widths of 3.33, 5.17 and so forth. For the results of psychological and educational measurements where fractional values do not occur, this can produce a histogram that is difficult to interpret. Also, note that in the graph above the axis labeled "Math" has the upper limits of the intervals given and that the interval 57–60 is considered by SPSS to start just above 57 (57.0000001) and go up to and include exactly 60, the highest score. This is different from what we get using "by hand" methods, but it will only occur when the highest value in the data corresponds to the upper limit of the highest interval and is a consequence of the particular way SPSS represents scores (which is different from the traditions of psychological and educational measurement).

We can use a similar procedure to provide a graphic representation of the cumulative frequency distribution. This graph, known as the **cumulative frequency curve** (also sometimes called an **ogive**), is prepared by placing values of the cumulative frequency on the ordinate, and scores or score intervals on the abscissa. A *point* representing the cumulative frequency for each score or interval is then plotted, and the points are connected, forming a graph such as the one shown in Figure 2–3. Note that the cumulative frequency curve never drops back toward the abscissa.

**Figure 2–3**
Cumulative frequency curve.




## MAKING THE COMPUTER DO IT

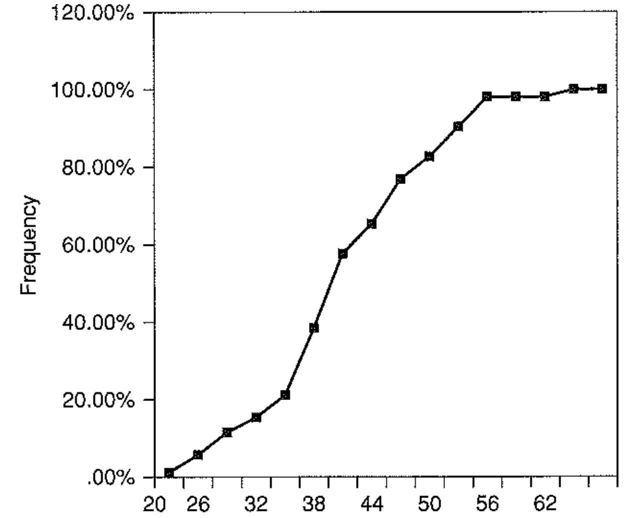
### Cumulative Frequency Curves

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.)

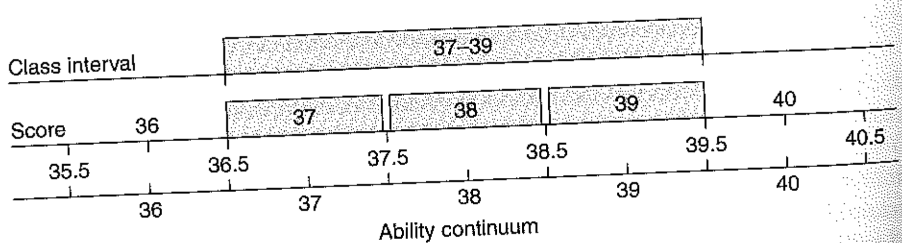



**Figure 2–5**
Relation between scores and ability continuum.



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 2/5ths 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

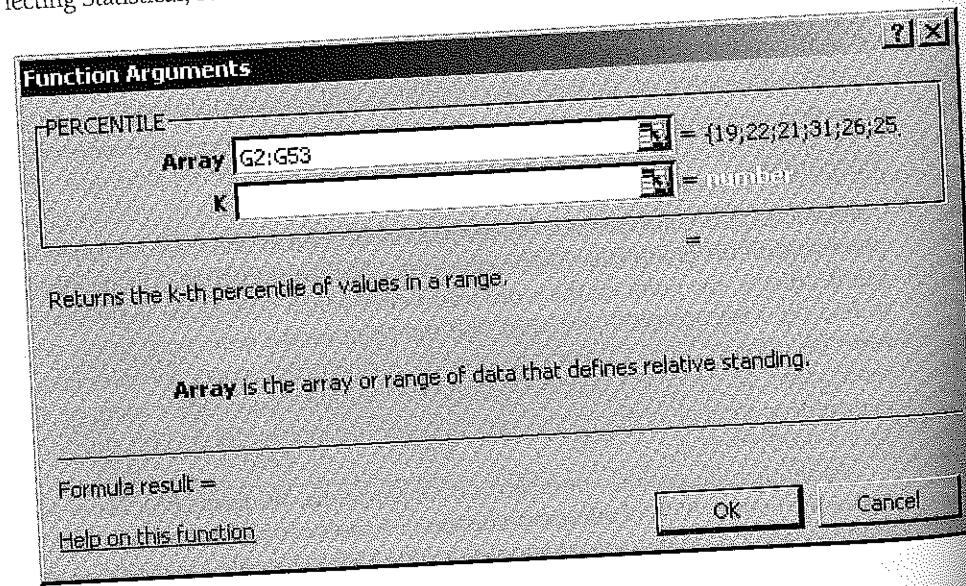
$$35.5 + 1.8 = 37.3$$

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 $\Sigma X$ 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

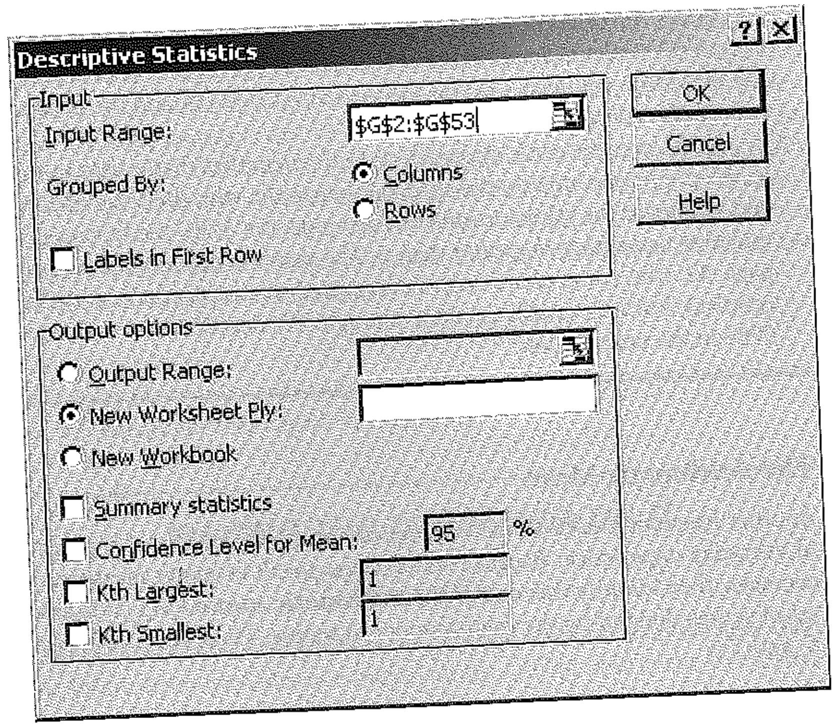
$$M = \frac{\Sigma X}{N} = \frac{1{,}985}{52} = 38.17$$

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.

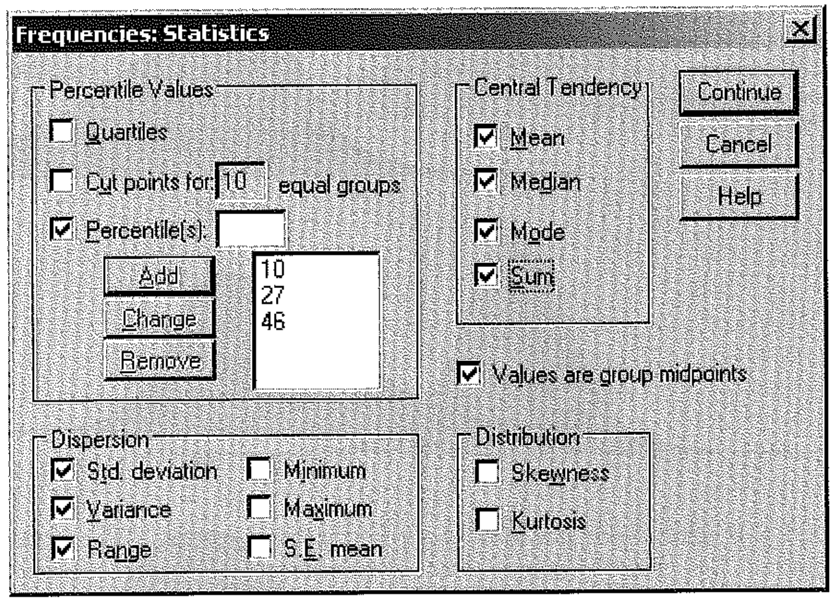



Click in the Input Range window on the pop-up box, then on the main screen highlight the scores to be included in computing the mean. If you are using the data from Table 2–1, the math test scores are in column G, rows 2–53 as we show here. Next click in the "Output Range" window of a pop-up box. Then 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.





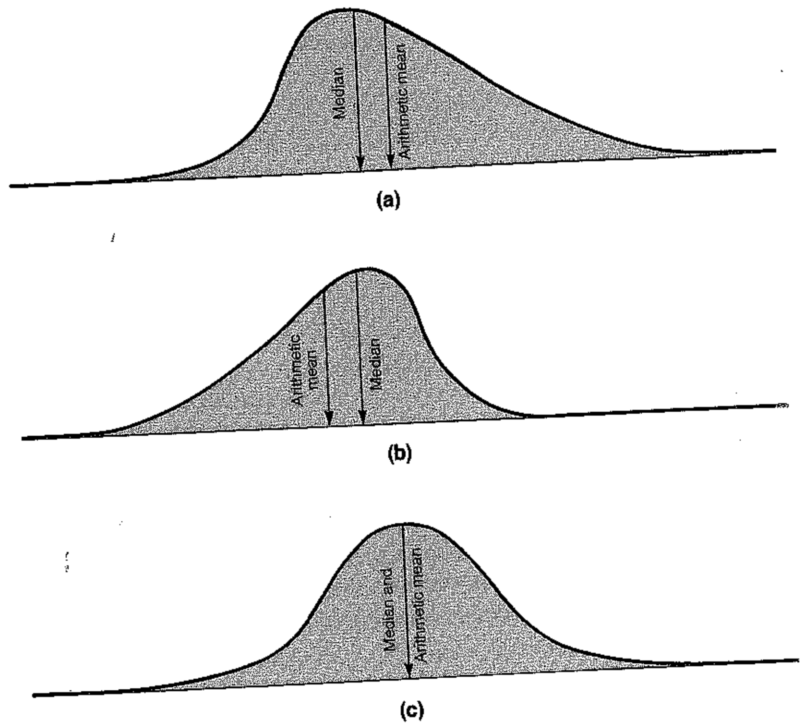

## Central Tendency and the Shape of the Distribution

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

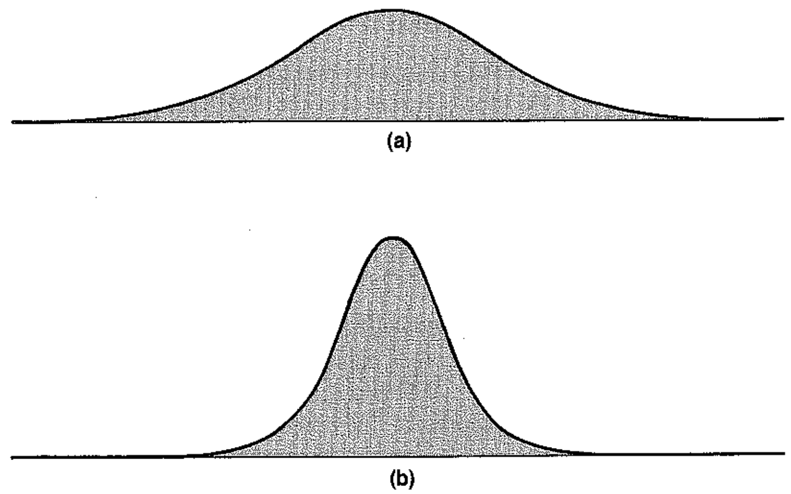
A simple measure of variability is the **range** of scores in the group, which is simply the difference between the highest and the lowest scores. In our math test example, the spread of scores is from 60 to 19, giving a range of 41 points. However, the range depends only on the two extreme cases in 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

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

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

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

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:

$$SD = \sqrt{\frac{(-1.5)^2 + (-.5)^2 + (.5)^2 + (1.5)^2}{4}}$$

$$SD = \sqrt{\frac{2.25 + 0.25 + 0.25 + 2.25}{4}} = \sqrt{\frac{5}{4}}$$

$$SD = \sqrt{1.25} = 1.12$$



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:

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

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 $f_X$ (Σ) 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.

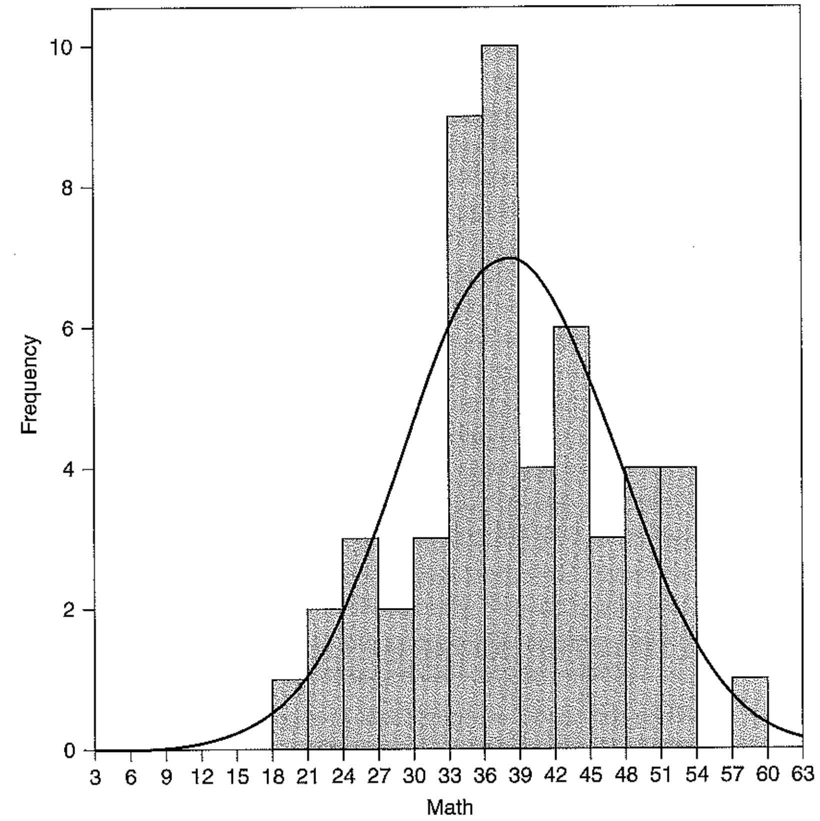
The standard deviation gets its most clear-cut meaning for one type of distribution of scores, the **normal distribution,** or **normal curve.** This distribution is defined by a particular mathematical equation, but to the everyday user, it is defined approximately by its pictorial qualities. The normal 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 *either* +1.0 *SD* or −1.0 *SD* | 34.1 |
| Between the mean and *either* +2.0 *SD* or −2.0 *SD* | 47.7 |
| Between the mean and *either* +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



of measures we have just been considering—percentiles and standard scores—provide a framework with which we can view the performance of a specific person. Both provide a way to view the person's performance relative to a specific *reference group.* Percentile information tells us what point in the score distribution just exceeds a specified fraction of the scores for the group; the standard deviation provides a common unit of distance from the mean of the group. The individual's score can then be expressed as a distance above or below the mean in these common units. In this way it is possible to give a meaningful answer to questions such as "Are you taller than you are heavy?"

## MEASURES OF RELATIONSHIP

We look now for a statistic to express the relationship between two sets of scores. For example, in Table 2–1, for each pupil we have scores for reading, mathematics, and spelling. To what extent did those pupils who scored well in mathematics also score well on the reading test? In this case, we have two scores for each individual. We can picture these scores using a graph in two dimensions, one dimension for each test. Such a graph is shown in Figure 2–9 and is called a **scatterplot.** The first person listed in Table 2–1, Aaron Andrews, had a reading test score of 32 and a mathematics test score of 43. His scores are represented by the * in Figure 2–9, plotted at 32 on the vertical, or reading, scale and at 43 on the horizontal, or mathematics, scale. The dots in the figure each represent one of the other 51 pupils' paired scores. For example, the dot at a reading score of 21 and a math score of 19 represents Quadra Quickly's scores.

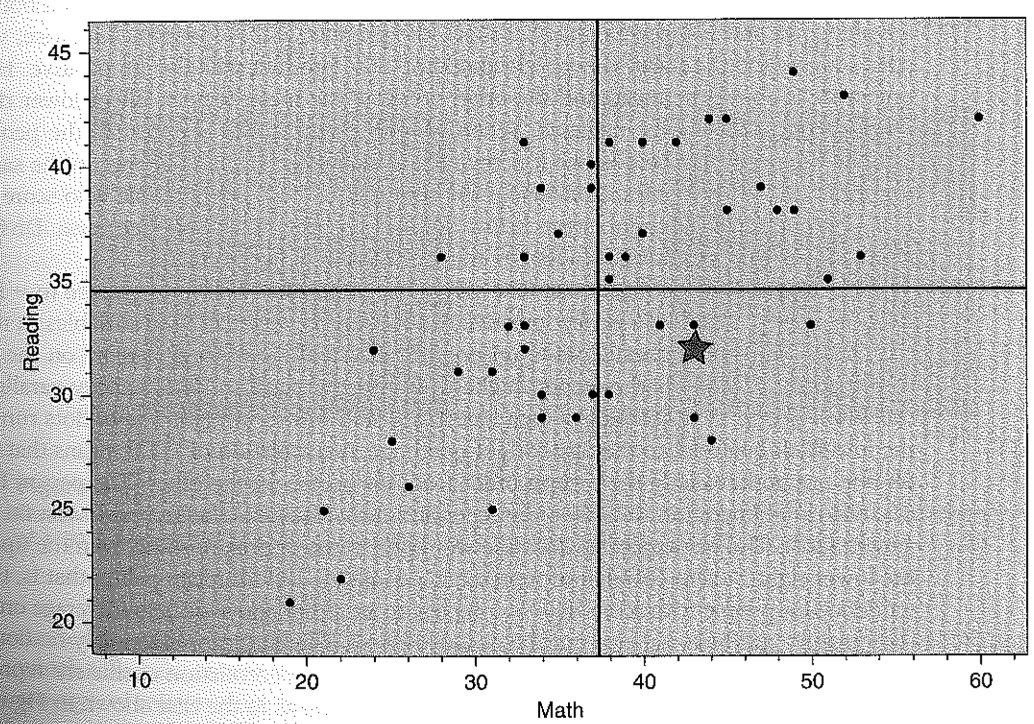



**Figure 2–9**
Scatterplot of reading and mathematics scores.

The vertical and horizontal lines drawn through the scatterplot are the means of the two variables and divide the plot into four parts, or **quadrants.** When a person who does well in reading (above the mean) also does well in mathematics, the dot representing that pair of scores falls in the upper right-hand quadrant. The dot for one who does poorly (below the mean) on both tests falls in the lower left quadrant. Where a good score on one test is paired with a poor score on the other, we find the points falling in the other quadrants, that is, the upper left and the lower right. Students who score in the middle on both tests are represented by points in the center of the plot. Inspection of Figure 2–9 reveals some tendency for the scores to scatter in the lower left to the upper right direction, from low reading *and* low mathematics to high reading *and* high mathematics, but there are many exceptions. The relationship is far from perfect; it is a matter of degree. We need some type of index to express this degree of relationship.

As an index of degree of relationship, a statistic known as the **correlation coefficient** is widely used. The symbol *r* is used to designate this coefficient. Looking at the formula for *r* will help us understand why the coefficient has the properties it does. The correlation coefficient is defined as

$$r = \frac{\Sigma Z_X Z_Y}{N}$$

where $Z_X$ and $Z_Y$ are the pair of standard scores for an individual (such as the +0.54 and −0.44 scores for Aaron Andrews). What the formula tells us to do is multiply each person's standard score on one variable by their standard score on the other, sum the products across all cases, and divide the result by the number of people in our sample.

Now think about those pairs of standard scores. If a person's Z-score is above the mean, it will have a positive sign; if it is below the mean, the sign will be negative. When both Z-scores are above the mean (upper right quadrant), the product will be positive. Likewise, when both scores are below the mean (lower left quadrant), the product will also be positive. However, if one score is above the mean and the other is below the mean (upper left or lower right quadrant), the product will be negative. When we add these products across all the people in our sample, the result will be positive if people tend, on average, to have scores on the same side of the mean on both variables. Conversely, the sum will be negative if people who score above the mean on one variable tend to score below the mean on the other, and vice versa. Dividing by *N* simply adjusts the resulting sum for the number of people in the group.

The correlation coefficient can take values ranging from +1.0 through zero to −1.0. A correlation of +1.0 signifies a perfect positive relationship between the two variables. It means that the person with the highest Z-score on one test also had the highest Z-score on the other, the next highest on one was the second highest on the other, and so forth, exactly parallel through the whole group. A scatterplot of data like this would form a straight line of dots running from the lower left quadrant to the upper right quadrant. A correlation of −1.0 means that the scores on one test go in exactly the reverse order from the scores on the other. The person highest on one test is the lowest on the other, the second highest on one is the second lowest on the other, and so forth. The scatterplot in this case would be a line of dots running from the upper left to the lower right. A zero correlation represents a complete lack of relationship; that is, there is no tendency for people who score high on one test to be either above or below average on the other (the positive products balance out the negative ones). The pattern is essentially random and the scatterplot will look like a circle. In-between values of the correlation coefficient represent tendencies for a relationship to exist, but with discrepancies, as in the case shown in Figure 2–9.

Every correlation coefficient contains two pieces of information. One is the *sign* of the correlation, which tells whether the two variables tend to rank people in the same order (plus) or in



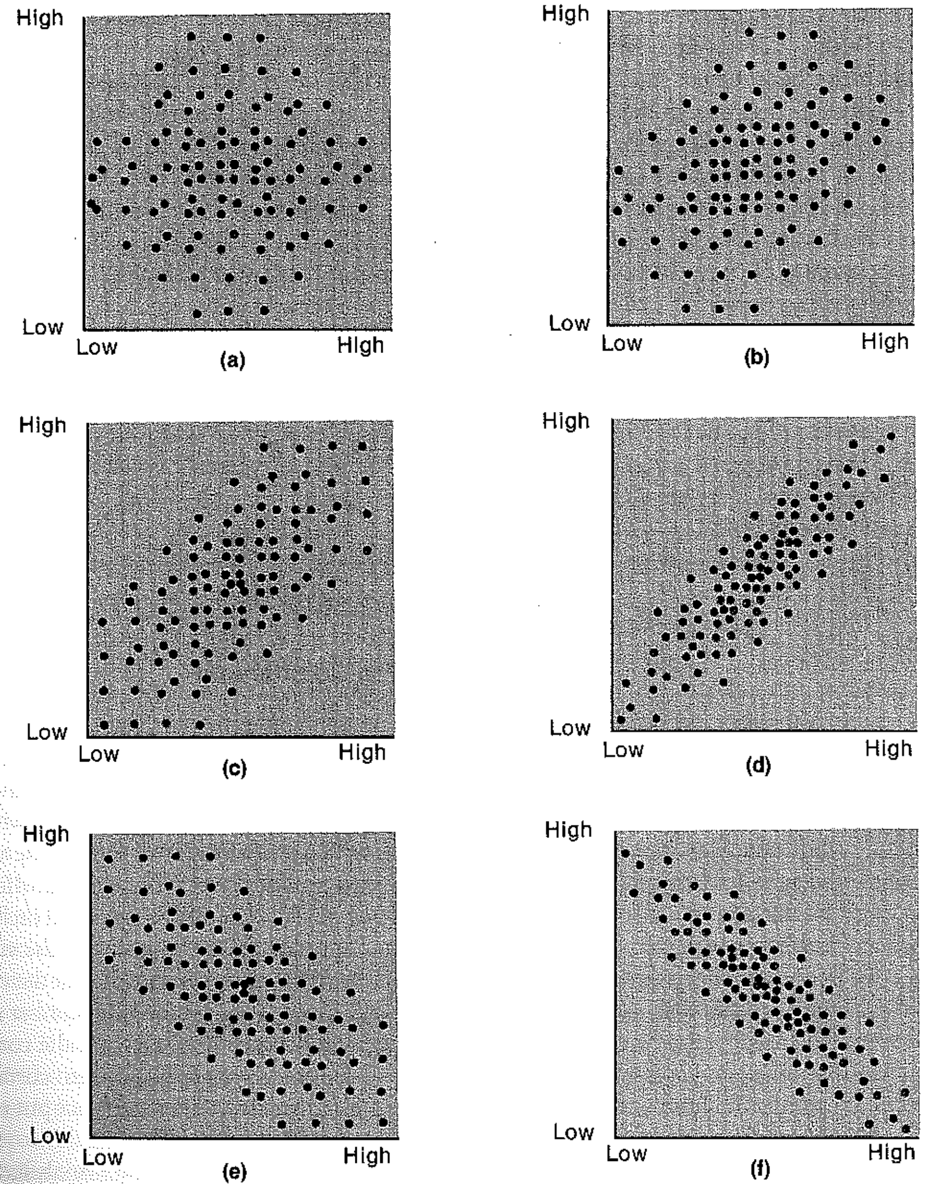



**Figure 2–10**
Distribution of scores for representative values of correlation coefficient: (a) correlation of 0.00, (b) correlation of +.30, (c) correlation of +.60, (d) correlation of +.90, (e) correlation of −.60, and (f) correlation of −.90.

the reverse order (minus). The second piece of information is the *magnitude* of the correlation, which tells how strong the relationship is. The correlations +.50 and −.50 indicate the same strength of relationship, but the first reveals that there is some tendency for people to be in the same rank order on both variables; the second shows some tendency for people with higher scores on one variable to have lower scores on the other. Figure 2–10 illustrates four different levels of relationships. In Figure 2–10a, the correlation is zero, and the points scatter in a pattern that is almost circular. All combinations are found equally: high–high, low–low, high–low, and low–high. Figure 2–10b corresponds to a correlation of +.30. You can see a slight trend for the points to group in the low–low and high–high direction. This tendency is more marked in Figure 2–10c, which represents a correlation of +.60. In Figure 2–10d, which portrays a correlation of +.90, the trend is much more pronounced. Note that when the correlation is −.60 (Figure 2–10e) or −.90 (Figure 2–10f), the scattering of the points is the same, but the swarm of dots falls along the opposite diagonal from that of Figures 2–10c and 2–10d—from the upper left-hand corner to the lower right-hand corner. But even with as high a correlation as +.90, the scores spread out quite a bit and do not all lie directly on the line from low–low to high–high. The scores plotted in Figure 2–9 correspond to a correlation coefficient of +.62, which is fairly high for most relationships between variables found in education and psychology.

As is the case for the standard deviation, many modern pocket calculators include a program to compute the correlation coefficient. Likewise, spreadsheet and statistics packages will always have programs to compute correlations, although they may be listed under the heading of *regression*, a closely related topic that we discuss next.

## MAKING THE COMPUTER DO IT

### Correlation Coefficients

Both SPSS and Excel have convenient routines for computing the correlations among several variables at the same time. SPSS has a Correlate program under the Analyze menu. Clicking on this program, you will be asked which of three types of correlation you wish to compute. Select "Bivariate," meaning two variables at a time. You will then be presented with a dialogue box that allows you to specify the variables to be correlated. Move each of the variables you want to find correlations for into the Variables box. The Pearson correlation, which is the one we have been discussing, is the one you want, so click OK. The output will include a square table in which each entry is the correlation of the variable in the row with the variable in the column. The values in the diagonal of the table will be 1.0, which says that the correlation of any variable with itself is perfect. Each cell of the table also includes the number of cases used to compute the correlation and the statistical significance ("Sig") test for the correlation (the probability that this correlation represents a random deviation from a zero correlation in the population from which this sample was drawn). Most investigators will not put much faith in a correlation where the value of Sig is larger than .05.

In Excel you simply select the Correlation option from the Data Analysis menu. A dialogue box will appear asking you for the Input Range. You must highlight all of the scores to be included in the analysis, or specify the upper left cell and the lower right cell of the part of the table containing the scores you want to analyze. Then click on the "Output Range" button in the dialogue box, click in the "Output Range" box, and click in an empty cell where you wish the upper left corner of the correlation table to go. The output will be the lower half of the table you would get from SPSS, including the ones in the diagonal. Excel also has a Correl function on the Σ or $f_X$ menu that you can use to find the correlation between a pair of variables. Highlight the scores on one variable for Array 1, then highlight the scores on the other variable for Array 2. Click OK, and the correlation will appear in the cell into which you have pasted the function.



You will encounter correlation coefficients in connection with testing and measurement in three important settings. The first situation is one in which we are trying to determine how precise and consistent a measurement procedure is. Thus, if we want to know how consistent a measure of speed we can expect to get from runners doing a 50-meter dash, we can have each person run the distance twice, perhaps on successive days. Correlating the two sets of scores will give information on the stability, or **reliability,** of this measure of running speed. The second situation is one in which we are studying the relationship between two different measures to evaluate one as a ***predictor*** of the other. Thus, we might want to study a scholastic achievement test from high school as a predictor of college grades. The correlation of the test scores with grades would give an indication of the test's usefulness as a predictor. These two uses of the correlation coefficient will be described more fully in Chapters 4 and 5.

The third situation in which we encounter correlation coefficients is more purely descriptive. We often are interested in the relationships between variables, simply to understand better how behavior is organized. What correlations do we find between measures of verbal and quantitative abilities? How close is the relationship between interest in mechanical jobs and comprehension of mechanical devices? Is rate of physical development related to rate of intellectual development? Many research problems in human behavior can best—or perhaps only—be studied by observing relationships as they develop in a natural setting, and these relationships often are expressed with correlation coefficients.

We face the problem, in each case, of evaluating the correlation we obtain. Suppose the two sets of 50-meter dash scores yield a correlation of +.80. Is this satisfactory? Suppose the achievement test scores correlate +.60 with college grades. Should we be pleased or discouraged? (The wording of this question implies that we want correlations to be high. When a correlation expresses the reliability, or consistency, of a test, or its accuracy in predicting an outcome of interest to us, it is certainly true that the higher the correlation, the more pleased we are. In other contexts, however, "bigness" does not necessarily correspond to "goodness," and we may not have a preference on the size of a correlation—or may even prefer a low one.)

The answer to the third question lies in part in the plots of Figure 2–10. Clearly, the higher the correlation, the more closely scores on one variable agree with scores on the other. If we think of discrepancies away from the diagonal line, from low–low to high–high, as "errors," the errors become smaller as the correlation becomes larger. If we think of the *standard deviation of scores on test Y* as the errors we would make in predicting people's scores without test *X*, then the correlation coefficient between *X* and *Y* tells us how much our errors will be reduced by using test *X* as a predictor of variable *Y*. The square of the correlation coefficient ($r^2$) tells us how much our errors in estimating people's test *Y* scores are reduced using test *X* over what those errors would be without the test. But these discrepancies are still discouragingly large for even rather substantial correlation coefficients, for example, for those shown in Figure 2–10. We must always be aware of these discrepancies and realize that even with a correlation such as +.60 between achievement test scores and school grades (which is about as high as correlations between these measures usually get), there still will be a number of students whose school performance will differ a good deal from the best prediction we can make from the test. We will discuss these issues in more detail in Chapter 5.

However, everything is relative, and any given correlation coefficient must be interpreted in comparison with values that are commonly obtained. Table 2–6 contains a number of correlations that have been reported for different types of variables. The nature of the scores being correlated is described, and the coefficient is reported. An examination of this table will provide some initial background for interpreting correlation coefficients. The correlation coefficient will gradually take on added meaning as you encounter coefficients of different sizes in your reading and as you work with tests.

**Table 2–6**
Correlations Between Selected Variables

| Variable | Correlation Coefficient |
|---|---|
| Heights of identical twins | .95 |
| Intelligence test scores of identical twins | .88 |
| Reading test scores in Grade 3 and Grade 6 | .80 |
| Rank in high school class and teacher's rating of work habits | .73 |
| Height and weight in 10-year-olds | .60 |
| Arithmetic computation test score and nonverbal intelligence test (Grade 8) | .54 |
| Height of brothers (adjusted for age) | .50 |
| Intelligence test score and parents' occupational level | .30 |
| Strength of grip and running speed | .16 |
| Adult height and intelligence test score | .06 |
| Ratio of head length to width and intelligence test score | .01 |
| Armed Forces Qualification Test scores of recruits and number of school grades repeated | −.27 |
| Artist interest score and banker interest score | −.64 |

## MAKING PREDICTIONS

We have seen that the correlation coefficient tells us the degree of association between two variables and the direction of that relationship. We can use this information, along with the means and standard deviations of the predictor and criterion variables, to answer our seventh question, how can we make the best possible prediction about criterion performance for each person, based on his or her score on the predictor? To do this we use what is called a **regression equation.** The regression equation tells us what our best guess of a person's score on an outcome variable, known as the criterion, would be, given the person's score on the predictor.

To see how the regression equation works, let us look at the scatterplot of the relationship between scores on the reading test and scores on the spelling test for Mr. Cordero's and Ms. Johnson's sixth-grade classes. Suppose we wish to use the spelling test to predict what level of reading proficiency we might expect from Heloise Abelard, a student who is being considered for transfer into this school. The scatterplot for the combined classes is shown in Figure 2–11. We have drawn two lines through the scatterplot. One is a horizontal line that shows the mean reading test score for all students for whom we have test scores. This is our best guess about Heloise's likely reading level *if we know nothing else about her.* In general, the mean of the criterion scores will always be our best guess about a person's score if we don't have any additional information. The regression equation is designed to allow us to make optimal use of any additional information we may have. What the regression equation does is describe the other line through the scatter plot, called the **regression line.** This line results in the most accurate predictions we can make *using the information from our predictor.*

We can use the regression line in either of two ways. The first is visual and is illustrated in Figure 2–12. Suppose Heloise earned a score of 50 on the spelling test. If we draw a line from the



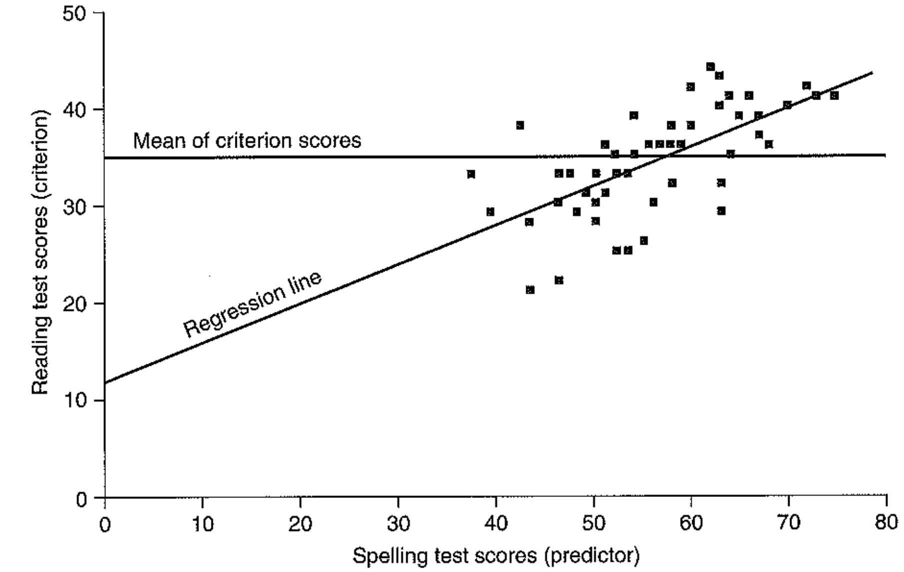



**Figure 2–11**
Scatterplot of scores from the reading and spelling tests given by Mr. Cordero and Ms. Johnson showing the overall mean of criterion scores and the regression line.

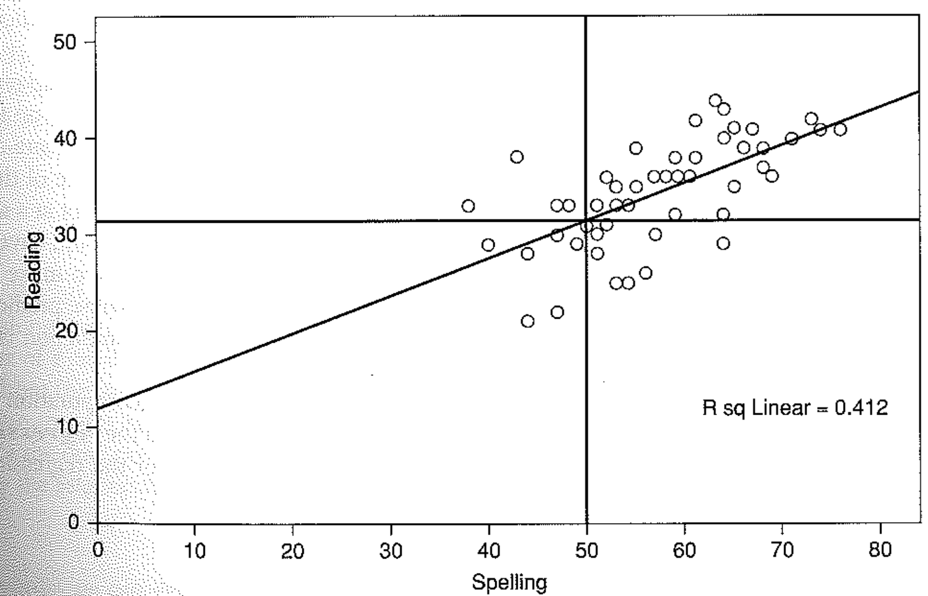



**Figure 2–12**
Using the regression line visually to make the best prediction of reading test score for Heloise.

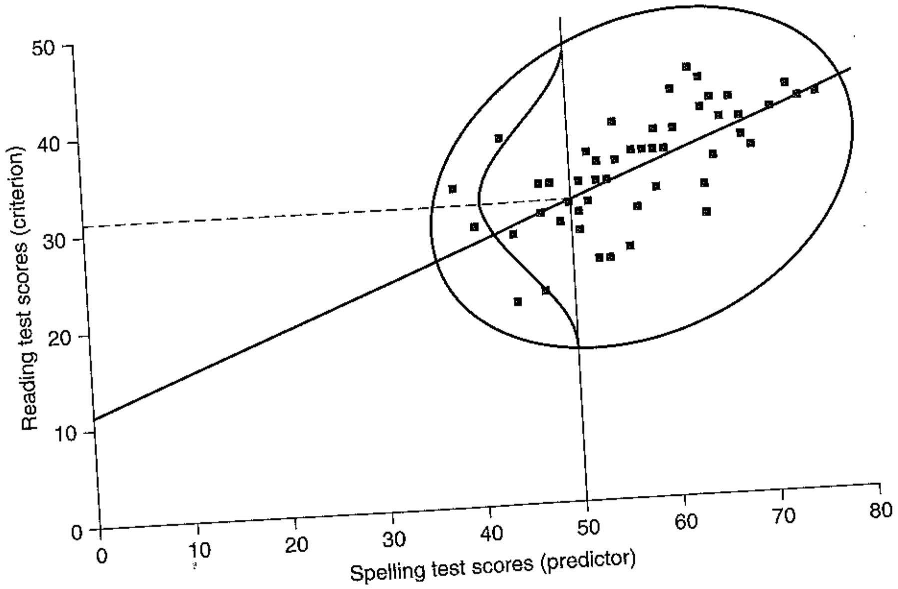



**Figure 2–13**
Scatterplot of reading and spelling test scores for a large group showing the distribution of reading test scores for all people who earned spelling scores of 50. The standard deviation of this distribution is the standard error of estimate.

spelling test score of 50 directly up to the regression line, and then horizontally over to the scale for the reading test, the value on the reading test scale will be the best prediction we can get of reading performance *for people with this spelling test score*. For Heloise, her predicted reading test score appears to be about 32.

If we had data on a very large number of people for which the relationship resembled that in Figure 2–12, many people would have spelling test scores of 50. The value on the reading test scale for the point on the regression line that corresponds to the spelling score of 50 would be the mean reading score for all people with that spelling score. This feature of the scatterplot is shown in Figure 2–13. Here a vertical slice is drawn through the scatterplot at a score of 50, and the normal distribution is the distribution of scores for all people who would have earned a score of 50 in the very large group. In this way, the regression line allows us to make predictions for new individuals. Our prediction is the mean of the theoretical group whose predictor score is 50, which is a point on the regression line, and the uncertainty of our prediction is shown by the spread of the distribution around the regression line.

The second way we can use the regression line is for direct computation of a predicted score. Any straight line can be described by a simple equation, and the ***regression equation*** is the equation that describes the regression line. The equation for the regression line has the general form

$$\hat{Y} = B_{YX}X + A$$

where $\hat{Y}$ is the predicted score on the criterion,

$B_{YX}$ is the *slope of the regression line* for predicting $Y$ from $X$ (the slope is the number of units that the line rises or falls in the scale of the criterion variable for every unit of increase we have in the predictor),



$X$ is the person's score on the predictor, and

$A$ is the *intercept*. The intercept is the value where the regression line crosses the criterion scale axis in the scatterplot. It is the value of the criterion we would predict for someone with a predictor score of zero.

Using relationships we will describe shortly, the slope and intercept for predicting reading score from spelling score are $B = .393$ and $A = 11.8$, respectively. Therefore, the predicted reading score for Heloise is

$$\hat{Y}_{Heloise} = .393(50) + 11.8 = 31.45$$

This is very close to the predicted score of 32 we got from examining the regression line in the scatterplot.

## Finding the Regression Line

An intimate relationship exists between the coefficients that define the regression line and the correlation coefficient. In fact, if we prepare a scatterplot of standard scores (Z-scores or scores in any other metric where *both variables have the same standard deviation*), the slope of the regression line *is* the correlation coefficient. That is, when $SD_Y = SD_X$,

$$B_{YX} = r_{YX}$$

However, when the two standard deviations are not equal, $B$ is related to $r$ by the ratio of the $SD$s. Specifically, when we are predicting $Y$ from $X$, the relationship is

$$B_{YX} = \left(\frac{SD_Y}{SD_X}\right) r_{YX}$$

To see how this equation works, let's take the data for our two sixth-grade classes and compute the slope coefficient for our regression equation. The standard deviations for the two variables are $SD_{Spelling} = 9.04$ and $SD_{Reading} = 5.55$. The correlation between the two variables is $r_{YX} = .64$, so the equation for $B_{YX}$ produces

$$B_{YX} = \left(\frac{5.55}{9.04}\right)(.64) = (.61)(.64) = .393$$

The value of $A$, the intercept, depends on the value of $B$ and the means of the two variables. Recalling that the intercept is the point on the criterion variable ($Y$) axis when $X$ is zero (or the value of $Y$ on the regression line where the regression line crosses the $Y$ axis), we can see that to get from the point where $X$ is zero to the point where $X = M_X$, we must increase the value of $X$ by the amount $M_X$. But if we increase $X$ by $M_X$, the regression line is going to change by $B(M_X)$. Since the only point we know for certain in the $Y$ distribution is the mean of $Y$ ($M_Y$), we can find $A$ by the relationship

$$A = M_Y - B_{YX}M_X$$

In words, we can say that the intercept is the mean of the criterion variable minus ($B$ times the mean of the predictor variable). For the two variables in our example we have $M_{Reading} = 34.44$ and $M_{Spelling} = 57.54$. The equation for $A$ therefore produces a value of

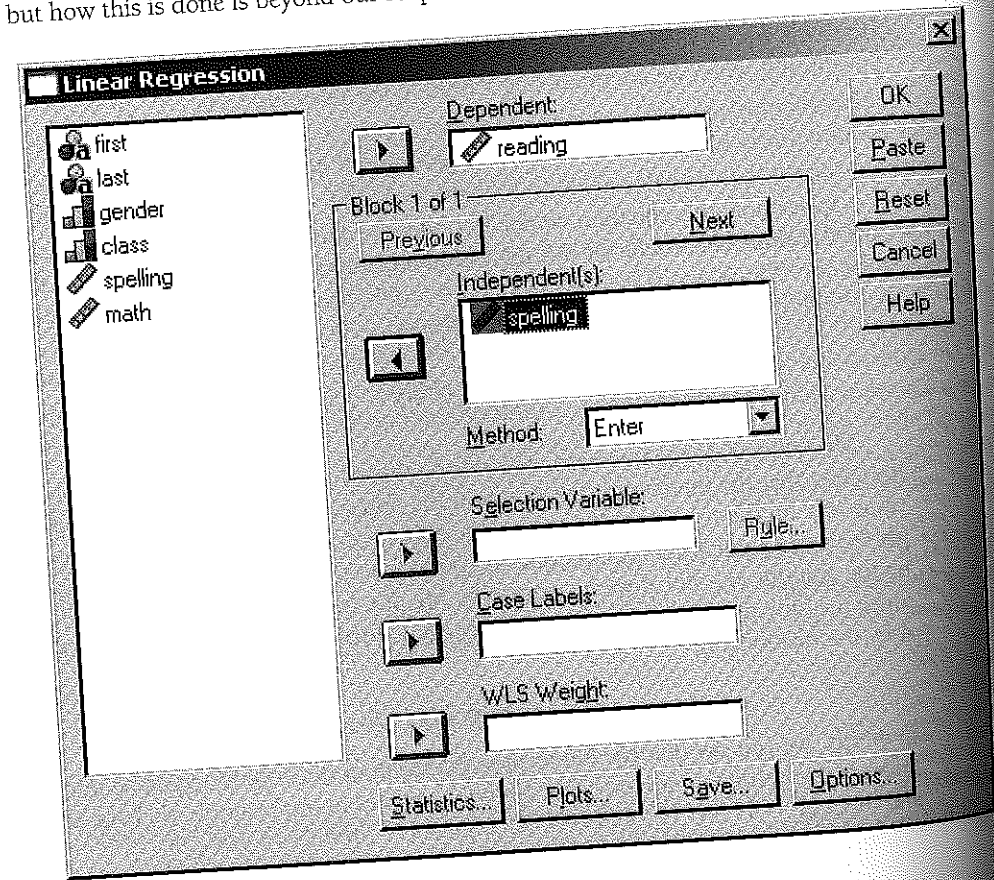
$$A = 34.44 - [(.393)(57.54)] = 11.8$$

which agrees with the graph in Figure 2–12. The equation $\hat{Y} = .393X + 11.8$ can be used to make predictions for as many new students as we wish. The same equation would be used for all new students until we have collected criterion data on them, at which point we might add them to our sample and recompute the values for the regression equation.

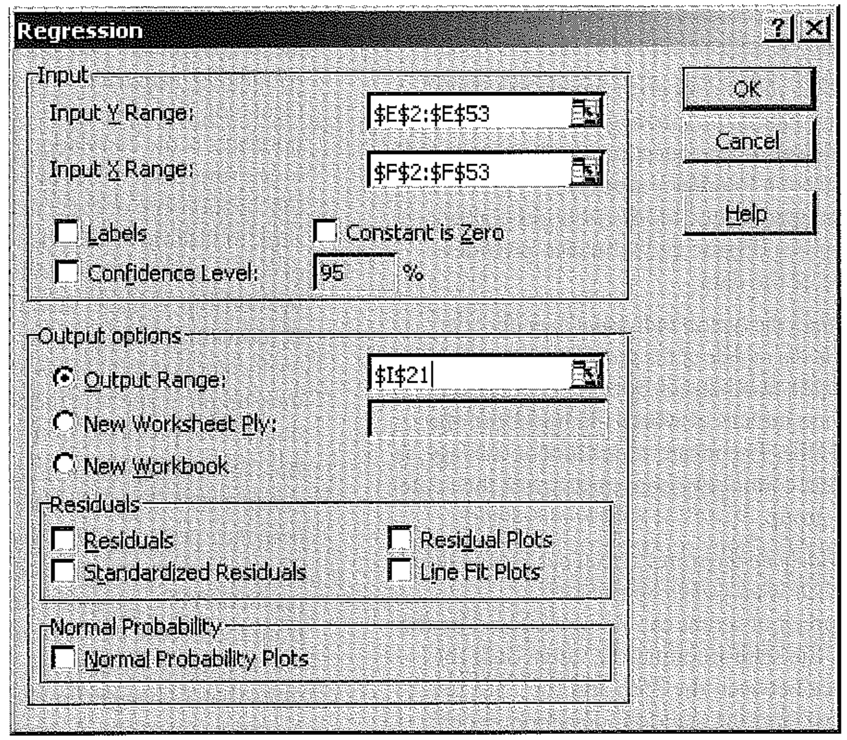
## MAKING THE COMPUTER DO IT

### Regression

It is quite simple to get the regression equation from either SPSS or Excel. Both have a program called Regression that will compute the necessary values. In SPSS the Regression program is under the Analysis menu. The program offers several types of regression, but the one we want is called "Linear." When you see the following screen, simply place the criterion variable in the Dependent box and the predictor in the Independent box. You can use the Statistics button to request additional output such as the means and standard deviations (under Descriptives). Click OK and you will get the regression slope and intercept. The value for the intercept is in the row labeled **Constant,** and the value for the predictor will be in the row with its label. The program can handle many predictors simultaneously, but how this is done is beyond our scope.






Excel has a regression program under the Data Analysis option on the Tools menu. When you select this option, you will be prompted to "Input Y range" and "Input X range." $Y$ is the criterion and you specify the cells in which the $Y$ scores are located. Do the same for $X$. If you wish the results to be displayed in the same spreadsheet as the data, select Output Range and specify a cell in which the output should start, usually a cell to the right of the data. For our example, the screen should look like the screen above.

If you use either program to replicate the analysis we have described in this section, you will get slightly different values than we did when these were calculated by hand. Both programs output values of $B = .394$ and $A = 11.76$. The reason for the difference is that the programs keep more decimal places in their intermediate calculations. It will often be the case that hand calculations will differ slightly from the computer for this reason.

## SUMMARY

All measurements produce numbers. It is usually desirable to summarize the numbers that result from measurements so that we can both answer specific questions about the people we have measured and draw general conclusions. We have focused on seven basic types of questions.

1. *What type of scale does our measurement represent?*
Different uses of numbers allow us to do different things with the resulting data. In nominal scales, numbers substitute for labels. Ordinal scales are usually in the form of ranks. Interval

and ratio scales allow us to find means and standard deviations.

2. *What does the general array of numbers "look like"? Or how do they arrange themselves?* The answer to this question involves sorting the scores into a frequency distribution (Table 2–2) or a grouped frequency distribution (Table 2–3). Histograms (Figure 2–1) can also provide a picture of the data.

3. *What does the typical individual look like? Or where is the middle of the group?* The mode, median (or 50th percentile), and mean are indices of where the middle of the group falls. The mean, or arithmetic average, is the most commonly used measure of the center of the group.

4. *How widely do the scores spread out around their center? Or what is their spread?* The spread of scores can be represented by the range, the semi-interquartile range (half the distance between the 25th and the 75th percentiles), and the standard deviation, which is a measure of how far scores deviate, on the average, from the mean.

5. *How should we interpret the score of any individual?* We will discuss this topic in Chapter 3.

6. *To what extent do the scores from two measurements "go together"? Or what is the degree of relationship between the two sets of scores?* Generally, the most useful measurement of relationships is the correlation coefficient. This index runs from +1.0, indicating a perfect positive relationship or exact agreement, through zero, indicating no association between the sets of scores, to −1.00, indicating a perfect negative relationship, or exact disagreement, on the order of individuals on the two traits. The correlation coefficient is important as an index of stability of measurement and as a measure of how well one trait can be predicted from another.

7. *How can we make the most accurate possible prediction of a person's performance on one measurement from their performance on another measurement?* The regression line provides the most accurate prediction of a person's score on variable Y from their score on variable X. The regression line is defined by its slope, B, or the number of units that Y changes for every unit of change in X, and its intercept, A, or the value of Y when X is zero. The slope of the regression line has the same sign as the correlation coefficient, but is affected by the relative sizes of the two variables' standard deviations.

## QUESTIONS AND EXERCISES

1. For each of the following sets of scores, select the most suitable score interval and set up a form for tallying the scores:

| Test | Number of Cases | Range of Scores |
|---|---|---|
| Mathematics | 103 | 15–65 |
| Reading Comprehension | 60 | 60–140 |
| Interest Inventory | 582 | 65–248 |

2. In each of the following distributions, indicate the size of the score interval, the midpoints of the intervals shown, and the real limits of the intervals (i.e., the dividing points between them):

| a. | b. | c. |
|---|---|---|
| 4–7 | 17–19 | 60–69 |
| 8–11 | 20–22 | 70–79 |
| 12–15 | 23–25 | 80–89 |
| . | . | . |
| . | . | . |

3. Using the spelling scores given in Table 2–1, make a frequency distribution and a histogram. Compute the median and the upper and lower quartiles. Also, compute the arithmetic mean and the standard deviation from the original scores.



4. The Bureau of Census uses the median in reporting average income. Why is this index used rather than the mean?

5. A 50-item mathematics test was given to the 150 students in five classes at Sunnyside School. Scores ranged from 16 to 50, with 93 of the students getting scores above 40. What would this score distribution look like? What could you say about the suitability of this test for this group? What measures of central tendency and variability would be most suitable? Why?

6. A high school teacher gave two sections of a biology class the same test. The results were as follows:

| | Section A | Section B |
|---|---|---|
| Median | 74.6 | 74.3 |
| Mean | 75.0 | 73.2 |
| 75th percentile | 79.0 | 80.0 |
| 25th percentile | 71.0 | 64.4 |
| Standard deviation | 6.0 | 10.5 |

From these data, what can you say about these two classes? What implications do the test results have for teaching the two groups?

7. A test in history given to 2500 10th-grade students had a mean of 52 and a standard deviation of 10.5. How many standard deviations above or below the mean would the following students fall?

| | | | | | |
|---|---|---|---|---|---|
| Heather | 48 | Rob | 60 | Marc | 31 |
| Krista | 56 | Tina | 36 | Bill | 84 |

8. If the distribution in Question 7 was approximately normal, what percentage of the group would Heather, Krista, Rob, Tina, Marc, and Bill each surpass?

9. Assuming that a set of scores is normally distributed with a mean of 82 and a standard deviation of 12, what would be the percentile rank for each of the following scores?
   a. 74
   b. 85
   c. 99

10. Explain the meaning of each of the following correlation coefficients:
    a. The correlation between scores on a reading test and scores on a group test of general intellectual ability is +0.78.
    b. Ratings of students on good citizenship and on aggressiveness show a correlation of −0.56.
    c. The correlation between weight and sociability is +0.02.

11. Jason's score on the Thorndike Precarious Prognosticator of Potential Proficiency ($TP^4$) is 87. The test has a mean of 100 and a standard deviation of 20. The mean freshman grade point average at Flunkemout University is 2.5 with a standard deviation of .5. The correlation between the $TP^4$ and grades at FU is −.60.
    a. What are the slope and intercept of the regression equation (be careful; watch the sign)?
    b. What is our best estimate of Jason's GPA at FU?

## SUGGESTED READINGS

Dretzke, B. J., & Heilman, K. A. (1998). *Statistics with Microsoft Excel.* Upper Saddle River, NJ: Prentice Hall.

Heiman, G. W. (2000). *Basic statistics for the behavioral sciences* (3rd ed.). Boston: Houghton Mifflin.

Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Fort Worth, TX: Harcourt Brace.

Runyon, R. P., Coleman, K. A., & Pittenger, D. J. (2000). *Fundamentals of behavioral statistics* (9th ed.). New York: McGraw-Hill.

Stevens, S. S. (1951). *Handbook of experimental psychology.* New York: Wiley.

Sweet, S. A. (1999). *Data analysis with SPSS.* Needham Heights, MA: Allyn & Bacon.