CHAPTER 14

DESCRIPTIVE STATISTICS

OBJECTIVES

After reading Chapter 14, you should be able to do the following:

1. List the steps involved in scoring standardized and self-developed tests.
2. Describe the process of coding data, and give three examples of variables that would require coding.
3. List the steps involved in constructing a frequency polygon.
4. Define or describe three measures of central tendency.
5. Define or describe three measures of variability.
6. List four characteristics of normal distributions.
7. List two characteristics of positively skewed distributions and negatively skewed distributions.
8. Define or describe two measures of relationship.
9. Define or describe four measures of relative position.
10. Generate (in other words, make up) a column of 10 numbers, each between 1 and 10. You may use any number more than once. Assume those numbers represent scores on a posttest. Using these "scores," give the formula and compute the following (show your work): mean, standard deviation, z scores, and Pearson r (divide the column in half and make two columns of five scores each).

The goal of chapters 14 and 15 is for you to be able to select, apply, and correctly interpret analyses appropriate for a given study. After you have read Chapter 15, you should be able to perform the following task.

TASK 9

Based on Tasks 2–5, which you have already completed, write the results section of a quantitative research report. Specifically,

1. Generate data for each of the participants in your study.
2. Summarize and describe data using descriptive statistics.
3. Statistically analyze data using inferential statistics.
4. Interpret the results in terms of your original research hypothesis.
5. Present the results of your data analyses in a summary table.

If SPSS is available to you, use it to check your work (see Task 9 Performance Criteria, p. 498).

THE WORD IS "STATISTICS," NOT "SADISTICS"

Statistics is a set of procedures for describing, synthesizing, analyzing, and interpreting quantitative data. For example, 1,000 scores can be represented by a single number. As another example, you would not expect two groups to perform exactly the same on a posttest, even if they were essentially equal. Application of the appropriate statistic helps you to decide if the difference between two groups' scores is big enough to represent a true rather than a chance difference.

Choice of appropriate statistical techniques is determined to a great extent by your research design, hypothesis, and the kind of data that will be collected. Thus, different research focuses lead to different statistical analyses. The statistical procedures and techniques of the study should be identified and described in detail in the research plan. Data analysis is as important as any other component of research. Regardless of how well the study is conducted, inappropriate analyses can lead to inappropriate research conclusions. Note, however, the complexity of the analysis is *not* necessarily an indication of its "goodness" or appropriateness.

There are many statistical approaches available to a researcher. This chapter and Chapter 15 will describe and explain those commonly used in educational research. The focus is on your ability to apply and interpret these statistics, not your ability to describe their theoretical rationale and mathematical derivation. Despite what you have heard, statistics is easy. To calculate the statistics in these chapters, you only need to know how to add, subtract, multiply, and divide. That is all. No matter how gross or complex a formula is, it can be turned into an arithmetic problem when applied to your data. The arithmetic problems involve only addition, subtraction, multiplication, and division; the formulas tell you how often, and in what order, to perform those operations. Even if you haven't had a math course since junior high school, you will be able to calculate statistics. The hardest formula requires arithmetic at the sixth-grade level. In fact, you are encouraged to use a calculator! All you have to do is follow the steps we present. Trust us. You are going to be pleasantly surprised to see just how easy statistics is.

PREPARING DATA FOR ANALYSIS

A research study usually produces a mass of raw data, such as the responses of participants to an achievement, ability, interest, or attitude test. Collected data must be accurately scored and systematically organized to facilitate data analysis.

Scoring Procedures

All instruments administered should be scored accurately and consistently; each participant's test results should be scored in the same way and with one criterion. When a standardized instrument is used, scoring is greatly facilitated. The test manual usually spells out the steps to follow in scoring each test, and a scoring key is usually provided. If the manual is followed conscientiously and each test is scored carefully, errors are minimized. It is usually a good idea to recheck all or at least some of the tests for consistency of scoring (say, 25% or every third test).

Scoring self-developed instruments is more complex, especially if open-ended items are involved. There is no manual to follow, and the researcher has to develop and refine a scoring procedure. Steps for scoring each item and for arriving at a total score must be delineated and carefully followed. If other than objective-type items (such as multiple-choice questions) are to be scored, it is advisable to have at least one other person independently score some or all of the tests as a reliability check. Planned scoring procedures should be tried out by administering the instrument to some individuals from the same or a similar population as the one from which research participants will be selected for the actual study. In this way, problems with the instrument or its scoring can be identified and corrected prior to the start of the study. The procedure ultimately used to score study data should be described in detail in the final research report.

Test questions that can be responded to on a standard, machine-scorable answer sheet can save a lot of time and increase the accuracy of the scoring process. If tests are to be machine scored, answer sheets should be checked carefully for stray pencil marks and a percentage of them should be scored by hand just to make sure that the key is correct and that the machine is scoring properly. The fact that the tests are being scored by a machine does not relieve the researcher of the responsibility of carefully checking data before and after processing.

Tabulation and Coding Procedures

After instruments have been scored, the results are transferred to summary data sheets, or more likely to a computer. Tabulation involves organizing the data. Recording the scores in a systematic manner facilitates examination and analysis of the data. If analysis consists of comparing the posttest scores of two or more groups, data would generally be placed in columns, one for each group, with the data arranged in ascending or descending order. If pretest scores are involved, additional columns should be formed. If analyses involve subgroup comparisons, scores should be tabulated separately for each subgroup. For example, in a study investigating the interaction between two types of mathematics instruction and two levels of aptitude (a 2 × 2 factorial design), four subgroups are involved, as shown in Table 14.1. This is the common method of dealing with quantitative data. If the data to be analyzed are categorical, tabulation usually involves counting responses. For example, a

TABLE 14.1 Hypothetical Results of a Study Based on a 2 × 2 Factorial Design

	METHOD A	METHOD B
High Aptitude	68	55
	72	60
	76	65
	78	70
	80	72
	84	74
	84	74
	85	75
	86	75
	86	76
	88	76
	90	76
	91	78
	92	82
	96	87
	METHOD A	METHOD B
Low Aptitude	50	60
	58	66
	60	67
	62	68
	64	69
	64	69
	65	70
	65	70
	66	71
	67	71
	70	72
	72	75
	72	76
	75	77
	78	79

superintendent might be interested in comparing the attitude toward unions of elementary and secondary teachers. Thus, for a question such as, "Would you join a union if given the opportunity?" the superintendent would tally the number of "yes," "no," and "undecided" responses separately for elementary and secondary teachers.

When a number of different kinds of data are collected from each participant, such as demographic information and several different test scores, both the variable names and the actual data are frequently coded. The variable "pretest reading comprehension scores," for example, may be coded as PRC, and gender of participants may be recorded as "M" or "F" or "1" or "2." Use of a computer for tabulating data and doing data analysis is recommended in general, but particularly if complex or multiple analyses are to be performed, or if a large number of participants are involved. In these cases, coding the data is especially important. The major advantage of using a computer to organize and analyze data is the capacity to rearrange data by subgroups and extract information without reentering all the data.

The first step in coding data is to give each participant an ID number. If there are 50 participants, for example, number them from 01 to 50. As this example illustrates, if the highest value for a variable is 2 digits (e.g., 50), then all represented values must be 2 digits. Thus, the first participant is 01, not 1. Similarly, achievement scores that range from 75 to 132 are coded 075 to 132. The next step is to make decisions as to how nonnumerical, or categorical, data will be coded. Nominal or categorical data include variables such as gender, group membership, and college level (e.g., sophomore). Thus, if the study involves 50 participants, with 2 groups of 25, then group membership may be coded "1" or "2" or "experimental" or "control." Categorical data also occur in survey instruments on which participants choose from a small number of alternatives representing a wider range of values. For example, teachers might be asked the following question:

> How many hours of classroom time do you spend per week in nonteaching activities?
> (a) 0–5 (b) 6–10 (c) 11–15 (d) 16–20

Responses might be coded (a) = 1, (b) = 2, (c) = 3, and (d) = 4.

Once the data have been prepared for analysis, the choice of statistical procedures to be applied is determined not only by the research hypothesis and design, but also by the type of measurement scale (categorical, ordinal, interval, ratio) represented by the data.

USING A COMPUTER

Generally the computer is a logical choice for data analysis. However, a good guideline for beginning researchers is, do not use the computer to perform an analysis that you have never done yourself by hand, or at least studied extensively. For example, after you have performed several analyses of variance on various sets of data, you will have the experience to understand the information produced by a computer analysis. In addition, instructions for preparing data for computer processing will make sense to you and you will know what the resulting output should look like.

Rapid advances and the development of "user friendly" computers have made it possible for researchers to perform a variety of analyses efficiently and accurately. One of the most popular statistical packages, commonly used in many colleges and universities, is the Statistical Package for the Social Sciences (SPSS). SPSS is used widely in quantitative research and is relatively easy to learn. We will demonstrate the usefulness of SPSS in this chapter and the next.

TYPES OF DESCRIPTIVE STATISTICS

The first step in data analysis is to describe, or summarize, the data using descriptive statistics. In some studies, particularly survey ones, the entire data analysis procedure may consist solely of calculating and interpreting descriptive statistics. **Descriptive statistics** permit the researcher to meaningfully describe many pieces of data with a few indices. If such indices are calculated for a sample drawn from a population, the resulting values are referred to as statistics; if they are calculated for an entire population, they are referred to as parameters. Most of the statistics used in educational research are based on data collected from well-defined samples, so most analyses deal with statistics, not parameters. Restated, a **statistic** is a quantitative index that describes performance of a sample or samples, and a **parameter** is a quantitative index describing the performance of a population.

The major types of descriptive statistics are measures of central tendency, measures of variability, measures of relative position, and measures of relationship. **Measures of central tendency** are used to determine the typical or average score of a group of scores. **Measures of variability** indicate how spread out a group of scores are. Measures of *relative position* describe a participant's performance compared to the performance of all other participants. Measures of *relationship* indicate the degree to which two sets of scores are related (remember correlation?). Before actually calculating any of these measures, it is often useful to present the data in graphic form.

GRAPHING DATA

As discussed, data are usually recorded on summary sheets or in computers—in columns, placed in ascending order. Data in this form are easily graphed, permitting the researcher to see what the distribution of scores looks like. The shape of the distribution may not be self-evident, especially if a large number of scores are involved, and, as we shall see later, the shape of the distribution may influence the researcher's choice of certain descriptive statistics.

The most common method of graphing data is to construct a frequency polygon. The first step is to list all scores and to tabulate how many subjects received each score. If 85 10th-grade students were administered an achievement test, the results might be as shown in Table 14.2.

TABLE 14.2 Frequency Distribution Based on 85 Hypothetical Achievement Test Scores

SCORE	FREQUENCY OF SCORE
78	1
79	4
80	5
81	7
82	7
83	9
84	9
85	12
86	10
87	7
88	6
89	3
90	4
91	1
	Total: 85 students

Once the scores are tallied, the steps are as follows:

1. Place all the scores on a horizontal axis, at equal intervals, from lowest score to highest.
2. Place the frequencies of scores at equal intervals on the vertical axis, starting with zero.
3. For each score, find the point where the score intersects with its frequency of occurrence and make a dot.
4. Connect all the dots with straight lines.

From Figure 14.1 we can see that most of the 10th graders scored at or near 85, with progressively fewer students achieving higher or lower scores. In other words, the scores appear to form a relatively normal or bell-shaped distribution, a concept we will discuss a little later. This knowledge would be helpful in selecting an appropriate measure of central tendency.

There are many types of other data-graphing approaches such as bar graphs, pie graphs, scatter plots (see Figure 11.1), box plots, and stem-and-leaf charts.[1] Examining a picture of the data can give some clues about which statistics are appropriate analyses.

MEASURES OF CENTRAL TENDENCY

Measures of central tendency provide a convenient way of describing a set of data with a single number. The number resulting from computation of a measure of central tendency represents the average or typical score attained by a group of subjects. The three most frequently encountered indices of central tendency are the mode, the median, and the mean. Each of these indices is used with a different scale of measurement: the mode is appropriate for describing nominal data, the median for describing ordinal data, and the mean for describing interval or ratio data. Since most quantitative measurement in educational research uses an interval scale, the mean is the most frequently used measure of central tendency.

The Mode

The **mode** is the score that is attained by more subjects than any other score. The data presented in Figure 14.1, for example, shows that the group mode is 85, since more participants (12) achieved that score than any other. The mode is not established through calculation; it is determined by looking at a set of scores or at a graph of scores and seeing which score occurs most frequently. There are several problems associated with the mode, and it is therefore of limited value and seldom used. For one thing, a set of scores may have two (or more) modes, in which case they are referred to as *bimodal*. Another problem with the mode is that it is an unstable measure of central tendency; equal-sized samples randomly selected from the same accessible population are likely to have different modes. However, when nominal data are being analyzed, the mode is the only appropriate measure of central tendency.

The Median

The **median** is that point, after scores are organized from low to high or high to low, above and below which are 50% of the scores. In other words, the median is the midpoint (like the median strip on a highway). If there are an odd number of scores, the median is the middle score (assuming the scores are arranged in order). For example, for the scores 75, 80, 82, 83, 87, the median is 82, because it is the middle score. If there is an even number of scores, the median is the point halfway between the two middle scores. For example, for the scores 21, 23, 24, 25, 26, 30, the median is 24.5; for the scores 50, 52, 55, 57, 59, 61, the median is 56.

[1]Wallgren, A., Wallgren, B., Persson, R., Jorner, U., and Haaland, J. (1996). *Graphing statistics and data*. Thousand Oaks, CA: Sage.

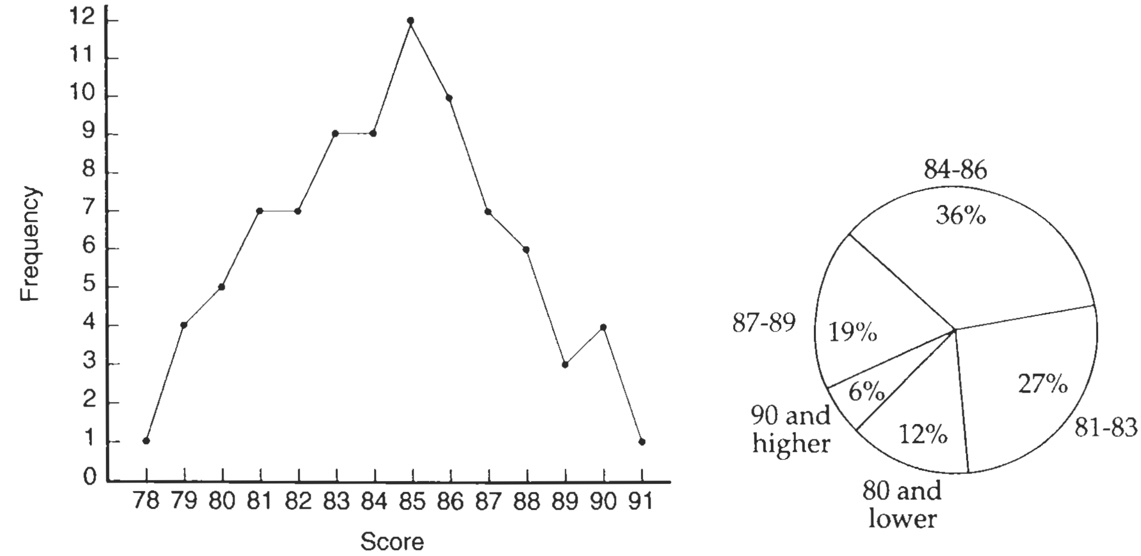


FIGURE 14.1
Frequency polygon and pie chart based on 85 hypothetical achievement test scores.

Thus, the median is not necessarily the same as one of the scores. There is no calculation for the median except finding the midpoint when there are an even number of scores.

The median does not take into account each and every score; it focuses on the middle scores. Two quite different sets of scores may have the same median. For example, for the scores 60, 62, 65, 67, 72, the median is 65; for the scores 30, 55, 65, 72, 89, the median is also 65. As we shall see shortly, this apparent lack of precision can be advantageous at times.

The median is the appropriate measure of central tendency when the data represent an ordinal scale. For certain distributions, the median may be the most appropriate measure of central tendency even though the data represent an interval or ratio scale. Although the median appears to be a rather simple index to determine, it cannot always be arrived at by simply looking at the scores; it does not always neatly fall between two different scores. For example, determining the median for the scores 80, 82, 84, 84, 84, 88 would require application of a relatively complex formula.

The Mean

The *mean* is the arithmetic average of the scores and is the most frequently used measure of central tendency. It is calculated by adding up all of the scores and dividing that total by the number of scores. In general, the mean is the preferred measure of central tendency. It is appropriate when the data represent either an interval or ratio score and is more precise than the median and the mode, because if equal-sized samples are randomly selected from the same population, the means of those samples will be more similar to each other than either the medians or the modes. By the very nature of the way in which it is computed, the mean takes into account, or is based on, each and every participant's score. Because all scores count, the mean can be affected by extreme scores. Thus, in certain cases, the median may actually give a more accurate estimate of the typical score.

When there are one or more extreme scores, the median will not be the most accurate representation of the performance of the total group but it will be the best index of typical performance. As an example, suppose you had the following IQ scores: 96, 96, 97, 99, 100, 101, 102, 104, 195. For these scores, the three measures of central tendency are

mode = 96 (most frequent score)
median = 100 (middle score)
mean = 110.6 (arithmetic average)

In this case, the median clearly best represents the typical score. The mode is too low, and the mean is higher than all of the scores except one. The mean is "pulled up" in the direction of the 195 score, whereas the median essentially ignores it. The different pictures presented by the different measures are part of the reason for the phrase, "lying with statistics." And in fact, selecting one index of central tendency over another one may present a particular point of view in a stronger light. In a labor-versus-management union dispute over salary, for example, very different estimates of typical employee salaries will be obtained depending on which index of central tendency is used. Let us say that the following are typical employee salaries in a union company: $12,000, $13,000, $13,000, $15,000, $16,000, $18,000, $45,000. For these salaries, the measures of central tendency are

mode = $13,000 (most frequent score)
median = $15,000 (middle salary)
mean = $18,857 (arithmetic average)

Both labor and management could overstate their case, labor by using the mode and management by using the mean. The mean is higher than every salary except one, $45,000, which in all likelihood would be the salary of a company manager. Thus, in this case, the most appropriate, and most accurate, index of typical salary would be the median. In research, we are not interested in "making cases" but rather in describing the data in the most accurate way. For the majority of sets of data the mean is the appropriate measure of central tendency.

Measures of Variability

Although measures of central tendency are very useful statistics for describing a set of data, they are not sufficient. Two sets of data that are very different can have identical means or medians. As an example, consider the following sets of data:

set A:	79	79	79	80	81	81	81
set B:	50	60	70	80	90	100	110

The mean of both sets of scores is 80 and the median of both is 80, but set A is very different from set B. In set A the scores are all very close together and clustered around the mean. In set B the scores are much more spread out; in other words, there is much more variation or variability in set B. Thus, there is a need for a measure that indicates how spread out the scores are, that is, how much variability there is. There are a number of descriptive statistics that serve this purpose, and they are referred to as measures of variability. The three most frequently encountered are the range, the quartile deviation, and the standard deviation. Although the standard deviation is by far the most often used, the range is the only appropriate measure of variability for nominal data, and the quartile deviation is the appropriate index of variability for ordinal data. As with measures of central tendency, measures of variability appropriate for nominal and ordinal data may be used with interval or ratio data even though the standard deviation is generally the preferred index for such data.

The Range

The **range** is simply the difference between the highest and the lowest score and is determined by subtraction. As an example, the range for the scores 79, 79, 79, 80, 81, 81, 81, is 2, while the range for the scores 50, 60, 70, 80, 90, 100, 110 is 60. Thus, if the range is small, the scores are close together; if it is large, the scores are more spread out. Like the mode, the range is not a very stable measure of variability, and its chief advantage is that it gives a quick, rough estimate of variability.

The Quartile Deviation

In "research talk" the **quartile deviation** is one half of the difference between the upper quartile and the lower quartile in a distribution. In English, the upper quartile is the 75th percentile, that point below which are 75% of the scores. Correspondingly, the lower quartile is the 25th percentile, that point below which are 25% of the scores. By subtracting the lower quartile from the upper quartile and then dividing the result by two, we get a measure of variability. If the quartile deviation is small, the scores are close together; if it is large, the scores are more spread out. The quartile deviation is a more stable measure of variability than the range and is appropriate whenever the median is appropriate. Calculation of the quartile deviation involves a process very similar to that used to calculate the median, which just happens to be the second quartile or the 50th percentile.

Variance

Variance indicates the amount of spread among test scores. If the variance is small, the scores are close together; if it is large, the scores are more spread out. The square root of the variance is called the *standard deviation* and, like variance, a small standard deviation indicates that scores are close together and a large one indicates that the scores are more spread out.

Calculation of the variance is quite simple. For example, five students took a test and received scores of 25, 25, 30, 40, and 30. The mean of these scores is—what? Right, 32. The difference of each student's score from the mean is

$35 - 32 = 3$
$25 - 32 = -7$
$30 - 32 = -2$
$40 - 32 = 8$
$30 - 32 = -2$ (Notice that the sum of the differences is 0. That's why we have to square the differences in the next step.)

Squaring each difference gives $9 + 49 + 4 + 64 + 2 = 130$. Dividing the squared differences by the number of scores gives us $130/5 = 26$. This is called the variance of the scores. Variance is seldom used itself, but is used to obtain the standard deviation. The standard deviation is the square root of the variance (26). Get your calculator out. The square root of 26 is 5.1, and this is the standard deviation of the five scores.

The Standard Deviation

The standard deviation is used when the data are interval or ratio, and is by far the most frequently used index of variability. Like the mean, its central tendency counterpart, the standard deviation is the most stable measure of variability and includes every score in its calculation. In fact, the first step in calculating the standard deviation is to find out how far away each score is from the mean by subtracting the mean from each score. If you know the mean and the standard deviation of a set of scores you have a pretty good picture of what the distribution looks like. If the distribution of scores is relatively normal or bell-shaped (about which we will have more to say shortly), then the mean plus 3 standard deviations and the mean minus 3 standard deviations encompass over 99% of the scores. In other words, each score distribution has its own mean and its own standard deviation that are calculated based on the scores. The number 3 is a constant. For any normal distribution of scores, the standard deviation multiplied by 3 and then added to the mean and subtracted from the mean will include almost all the scores in the distribution. The symbol for the mean is $\overline{X}$ and the standard deviation is usually abbreviated as *SD*. Thus, the concept described here can be expressed as follows: $\overline{X} \pm 3\,SD = 99+\%$ of the scores.

As an example, suppose that the mean of a set of scores ($\bar{X}$) is calculated to be 80 and the standard deviation (*SD*) to be 1. In this case the mean plus 3 standard deviations, $\bar{X} + 3\ SD$, is equal to $80 + 3(1) = 80 + 3 = 83$. The mean minus 3 standard deviations, $\bar{X} - 3\ SD$, is equal to $80 - 3(1) = 80 - 3 = 77$. Thus, almost all the scores fall between 77 and 83. This makes sense since, as we mentioned before, a small standard deviation (in this case $SD = 1$) indicates that the scores are close together, not very spread out.

As another example, suppose that a different set of scores had a mean ($\bar{X}$) calculated to be 80, but this time the standard deviation (*SD*) is calculated to be 4. In this case the mean plus three standard deviations, $\bar{X} + 3\ SD$, is equal to $80 + 3(4) = 80 + 12 = 92$. In case you still do not see,

$$80 + 1\ SD = 80 + 4 = 84$$
$$80 + 2\ SD = 80 + 4 + 4 = 88$$
$$80 + 3\ SD = 80 + 4 + 4 + 4 = 92$$

Or, to explain it another way, 80 plus 1 $SD = 80 + 4 = 84$, plus another $SD = 84 + 4 = 88$, plus one more (the third) $SD = 88 + 4 = 92$. Now, the mean minus three standard deviations, $\bar{X} - 3\ SD$, is equal to $80 - 3(4) = 80 - 12 = 68$. In other words,

$$80 - 1\ SD = 80 - 4 = 76$$
$$80 - 2\ SD = 80 - 4 - 4 = 72$$
$$80 - 3\ SD = 80 - 4 - 4 - 4 = 68$$

Or, to explain it another way, 80 minus 1 $SD = 80 - 4 = 76$, minus another $SD = 76 - 4 = 72$, minus one more (the third) $SD = 72 - 4 = 68$. Thus, almost all the scores fall between 68 and 92. This makes sense since a larger standard deviation (in this case $SD = 4$) indicates that the scores are more spread out. Clearly, if you know the mean and standard deviation of a set of scores, you have a pretty good idea of what the scores look like. You know the mean score and you know how spread out or variable the scores are. Using both, you can describe a set of data quite well.

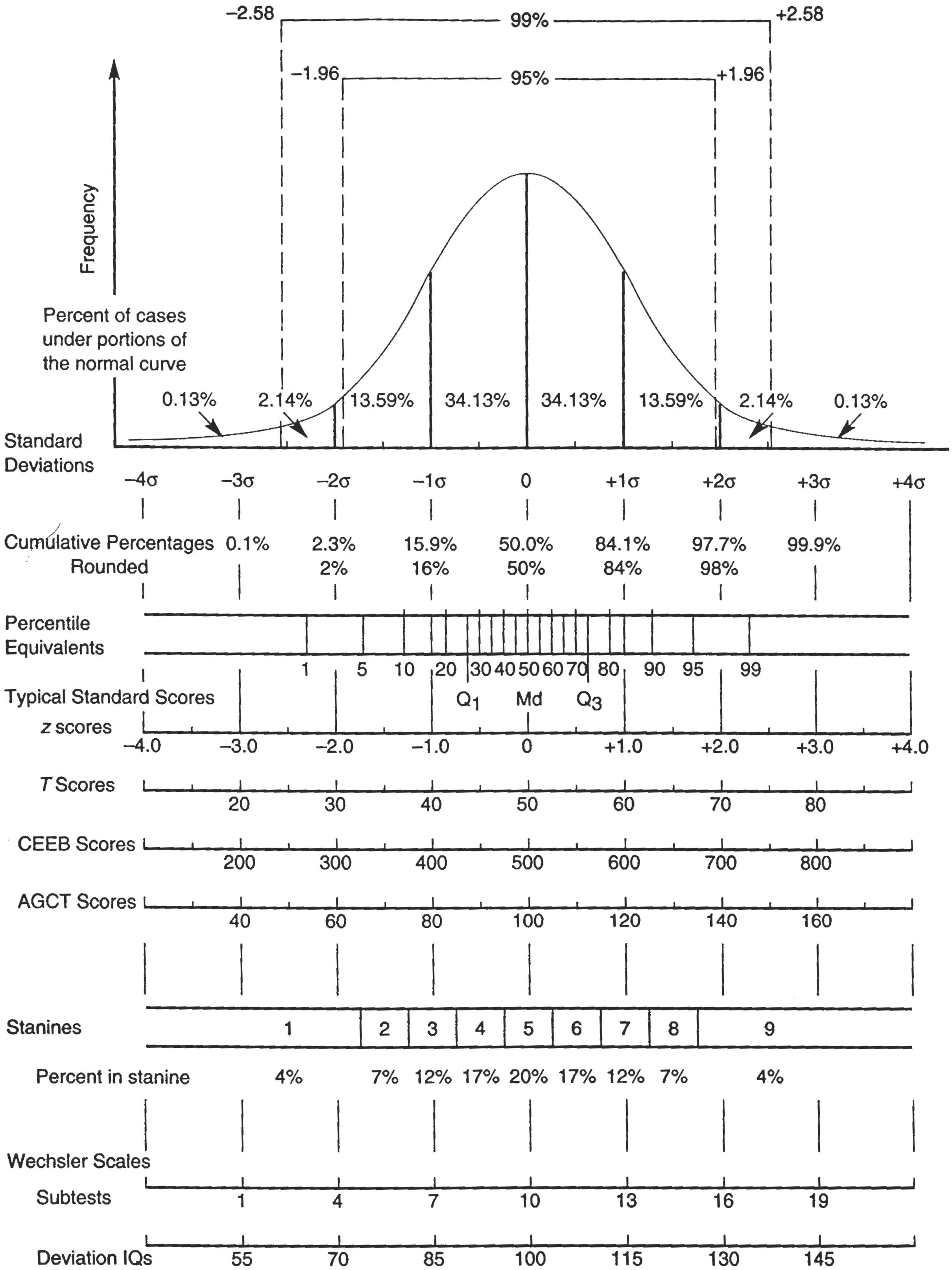
THE NORMAL CURVE

The plus and minus 3 concept is valid only when the scores are normally distributed, that is, form a normal, or bell-shaped, score distribution. Many, many variables, such as height, weight, IQ scores, and achievement scores yield a normal curve if a sufficient number of participants are measured.

If a variable is *normally distributed,* that is, forms a *normal* or *bell-shaped curve,* then several things are true:

1. Fifty percent of the scores are above the mean and 50% are below the mean.
2. The mean, the median, and the mode are the same value.
3. Most scores are near the mean and the farther from the mean a score is, the fewer the number of participants who attained that score.
4. The same number, or percentage, of scores is between the mean and plus one standard deviation ($\bar{X} + 1\ SD$) as is between the mean and minus one standard deviation ($\bar{X} - 1\ SD$), and similarly for $\bar{X} \pm 2\ SD$ and $\bar{X} \pm 3\ SD$ (Figure 14.2).

In Figure 14.2, the symbol σ (the Greek letter sigma) is used to represent the standard deviation, that is, $1\ \sigma = 1\ SD$, and the mean ($\bar{X}$) is designated as 0 (zero). The vertical lines at each of the *SD* (σ) points delineate a certain percentage of the total area under the curve. As Figure 14.2 indicates, if a set of scores forms a normal distribution, the $\bar{X} + 1\ SD$ includes



Note. This chart cannot be used to equate scores on one test to scores on another test. For example, both 600 on the CEEB and 120 on the AGCT are one standard deviation above their respective means, but they do not represent "equal" standings because the scores were obtained from different groups.

FIGURE 14.2 Characteristics of the normal curve (*Note:* Based on a figure appearing in *Test Service Bulletin* No. 48, January, 1955, of The Psychological Corporation.)

34.13% of the scores and the $\overline{X} - 1\ SD$ includes 34.13% of the scores. Each succeeding standard deviation encompasses a constant percentage of the cases. Since $\overline{X} \pm 2.58\ SD$ includes 99% of the cases, we see that $\overline{X} \pm 3\ SD$ includes almost all the scores, as pointed out previously.

Below the row of *SD*s in the figure is a row of percentages. As you move from left to right, from point to point, the cumulative percentage of scores that fall below each point is indicated. Thus, at the point that corresponds to $-3\ SD$, we see that only .1% of the scores fall below this point. The numerical value corresponding to $+1\ SD$, on the other hand, is a figure higher than 84.1% (rounded to 84% on the next row) of the scores. Relatedly, the next row, percentile equivalents, also involves cumulative percentages. The figure 20 in this row, for example, indicates that 20% of the scores fall below this point. While we will discuss percentiles and the remaining rows further as we proceed through this chapter, we will look at one more row at this time. Near the bottom of Figure 14.2, under Wechsler Scales, is a row labeled Deviation IQs. This row indicates that the mean IQ for the Wechsler Scale is 100 and the standard deviation is 15 (115 is in the column corresponding to $+1\ SD\ [+1\ \sigma]$) and since the mean is 100, 115 represents $\overline{X} + 1\ SD = 100 + 15 = 115$. An IQ of 145 represents a score 3 *SD*s above the mean (average) IQ. If your IQ is in this neighborhood, you are certainly a candidate for Mensa! An IQ of 145 corresponds to a percentile of 99.9. On the other side of the curve we see that an IQ of 85 corresponds to a score one standard deviation below the mean ($\overline{X} - 1\ SD = 100 - 15 = 85$) and to the 16th percentile. Note that the mean *always* corresponds to the 50th percentile. In other words, the average score is always that point above which are 50% of the cases and below which are 50% of the cases. Thus, if scores are normally distributed the following statements are true:

$\overline{X} \pm 1.0\ SD$ = approximately 68% of the scores
$\overline{X} \pm 2.0\ SD$ = approximately 95% of the scores (1.96 *SD* is exactly 95%)
$\overline{X} \pm 2.5\ SD$ = approximately 99% of the scores (2.58 *SD* is exactly 99%)
$\overline{X} \pm 3.0\ SD$ = approximately 99+% of the scores

And similarly, the following are always true:

$\overline{X} - 3.0\ SD$ = approximately the .1 percentile
$\overline{X} - 2.0\ SD$ = approximately the 2nd percentile
$\overline{X} - 1.0\ SD$ = approximately the 16th percentile
$\overline{X}$ = the 50th percentile
$\overline{X} + 1.0\ SD$ = approximately the 84th percentile
$\overline{X} + 2.0\ SD$ = approximately the 98th percentile
$\overline{X} + 3.0\ SD$ = approximately the 99th+ percentile

You may have noticed that the ends of the curve never touch the baseline and that there is no definite number of standard deviations that corresponds to 100%. This is because the curve allows for the existence of unexpected extremes at either end and because each additional standard deviation includes only a tiny fraction of a percent of the scores. As an example, for the IQ test the mean plus 5 standard deviations would be $100 + 5(15) = 100 + 75 = 175$. Surely 5 *SD*s would include everyone. Wrong! There has been a very small number of persons who have scored near 200, which corresponds to $+6.67$ *SD*s. Thus, while ± 3 *SD*s includes just about everyone, the exact number of standard deviations required to include every score varies from variable to variable.

As mentioned earlier, many variables form a normal distribution, including physical measures, such as height and weight, and psychological measures, such as intelligence and aptitude.

In fact, most variables measured in education form normal distributions if enough subjects are tested. Note, however, that a variable that is normally distributed in a population may not be normally distributed in smaller samples from the population. In Figure 14.2 the standard deviation is symbolized as σ, instead of *SD*, to indicate that the curve represents the scores of a population, not a sample; thus, σ represents a population parameter, whereas *SD* represents a sample-based statistic. Depending on the size and nature of a particular sample, the assumption of a normal curve may or may not be a valid one. Since research studies deal with a finite number of participants, and often not a very large number, research data only more or less approximate a normal curve. Correspondingly, all of the equivalencies (standard deviation, percentage of cases, and percentile) are also only approximations. This is an important point, since most statistics used in educational research are based on the assumption that the variable is normally distributed. If this assumption is badly violated in a given sample, then certain statistics should not be used. In general, however, the fact that most variables are normally distributed allows us to quickly determine many useful pieces of information concerning a set of data.

Skewed Distributions

When a distribution is not normal, it is said to be skewed. A normal distribution is symmetrical and the values of the mean, the median, and the mode are the same. A **skewed distribution** is not symmetrical, and the values of the mean, the median, and the mode are different. In a symmetrical distribution, there are approximately the same number of extreme scores (very high and very low) at each end of the distribution. In a skewed distribution there are more extreme scores at one end than the other. If the extreme scores are at the lower end of the distribution, the distribution is said to be **negatively skewed,** and if the extreme scores are at the higher end of the distribution, the distribution is said to be **positively skewed** (Figure 14.3).

As we can see by looking at the negatively skewed distribution, most of the participants did well but a few did very poorly. Conversely, in the positively skewed distribution, most of the participants did poorly but a few did very well. In both cases, the mean is "pulled" in the direction of the extreme scores. Since the mean is affected by extreme scores (all scores are used) and the median is not (only the middle score(s) are used), the mean is always closer to the extreme scores than the median. Thus, for a negatively skewed distribution the mean ($\overline{X}$) is always lower, or smaller, than the median (md); for a positively skewed distribution the mean is always higher or greater than the median. Since the mode is not affected by extreme scores, no "always" statements can be made concerning its relationship to the mean and the median in a skewed distribution. Usually, however, as Figure 14.3 indicates, in a negatively skewed distribution the mean and the median are lower, or smaller, than the mode, whereas in a positively skewed distribution the mean and the median are higher, or greater, than the mode.

To summarize:

negatively skewed: mean $<$ median $<$ mode
positively skewed: mean $>$ median $>$ mode

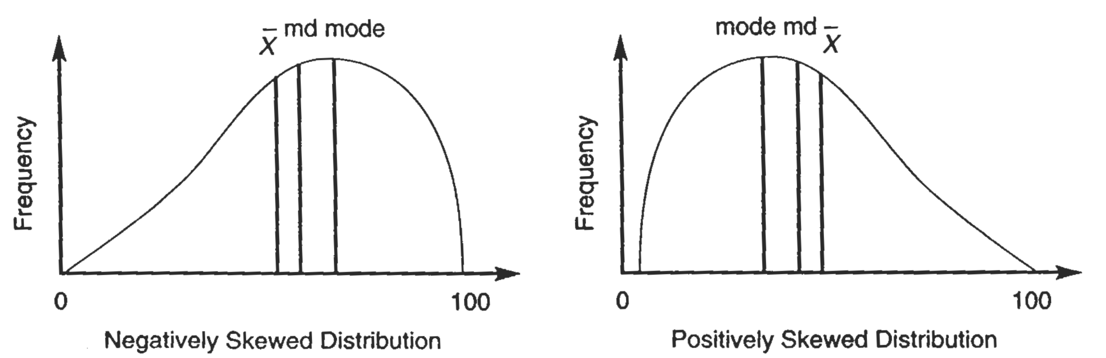


FIGURE 14.3 A positively skewed distribution and a negatively skewed distribution, each resulting from the administration of a 100-item test.
(*Note.* $\overline{X}$ = mean; md = median)

Because the relationship between the mean and the median is a constant, the skewness of a distribution can be determined without constructing a frequency polygon. If the mean is less than the median, the distribution is negatively skewed; if the mean and the median are the same, or very close, the distribution is symmetrical; if the mean is greater than the median, the distribution is positively skewed. The farther apart the mean and the median are, the more skewed is the distribution. If the distribution is very skewed, then the assumption of normality required for many statistics is violated.

MEASURES OF RELATIVE POSITION

Measures of relative position indicate where a score is in relation to all other scores in the distribution. In other words, measures of relative position permit one to express how well an individual has performed as compared to all other individuals in the sample who have been measured on the same variable. In Chapter 5 this was called *norm-referenced measurement.* A major advantage of such measures is that they make it possible to compare the performance of an individual on two or more different tests. For example, if Ziggy's score in reading is 40 and his score in math is 35, it does not follow that he did better in reading; 40 may have been the lowest score on the reading test and 35 the highest score on the math test! Measures of relative position express different scores on a common scale. The two most frequently used measures of relative position are percentile ranks and standard scores.

Percentile Ranks

A *percentile rank* indicates the percentage of scores that fall at or below a given score. If Matt Mathphobia's score of 65 corresponds to a percentile rank of 80, the 80th percentile, this means that 80 percent of the scores in the distribution are lower than 65. Matt scored higher than 80 percent of those taking the test. Conversely, if Dudley Veridull scored at the 7th percentile, this would mean that Dudley only did better than, or received a higher score than, 7% of the test takers.

Percentiles are appropriate for data representing an ordinal scale, although they are mainly computed for interval data. The median of a set of scores corresponds to the 50th percentile, which makes sense since the median is the middle point and therefore the point below which are 50% of the scores. While percentile ranks are not used very often in research studies, they are frequently used in the public schools to report test results of students in a form that is understandable to most audiences.

Standard Scores

Figure 14.2 depicts a number of standard scores. Basically, a *standard score* is a derived score that expresses how far a given raw score is from some reference point, typically the mean, in terms of standard deviation units. A standard score is a measure of relative position which is appropriate when the test data represent an interval or ratio scale of measurement. The most commonly reported and used standard scores are z scores, T scores (or Z scores), and stanines. Standard scores allow scores from different tests to be compared on a common scale and, unlike percentiles, permit valid mathematical operations such as averages to be computed. Averaging nonstandard scores on a series of tests in order to obtain an overall average score is like averaging apples and oranges and getting an "orapple." Such tests are likely to vary in level of difficulty and variability of scores. By converting test scores to standard scores, however, we can average them and arrive at a valid final index of average performance.

The normal curve equivalencies indicated in Figure 14.2 for the various standard scores are accurate only to the degree to which the distribution is normal. Further, standard scores can be compared only if all the derived scores are based on the raw scores (number correct) of the same group. For example, a CEEB (College Entrance Examination

Board) score of 700 is not equivalent to a Wechsler IQ of 130 because the tests were normed on different groups. If a set of raw scores is normally distributed, then so are the standard score equivalents. But, as noted, all distributions are not normal. For example, height is normally distributed, but the measured heights of the girls in a seventh-grade gym class may not be. There is a procedure for transforming raw scores that ensures that the distribution of standard scores will be normal. Raw scores thus transformed are referred to as *normalized* scores. All resulting standard scores are normally distributed and the normal curve equivalencies are accurate.

z **Scores.** A ***z* score** is the most basic standard score. It expresses how far a score is from the mean in terms of standard deviation units. A score that is exactly "on" the mean corresponds to a *z* score of 0. A score that is exactly 1 standard deviation above the mean corresponds to a *z* score of +1.00 and a *z* score that is exactly 2 standard deviations below the mean corresponds to a *z* score of −2.00. Get it? As Figure 14.2 indicates, if a set of scores is transformed into a set of *z* scores (each score is expressed as a *z* score), the new distribution has a mean of 0 and a standard deviation of 1.

The major advantage of *z* scores is that they allow scores from different tests or subtests to be compared. As an example, suppose Bobby Bonker's mother, a woman who is really on top of things, comes in and asks his teacher, "How is Bobby doing in the basic skills area?" If the teacher tells her that Bobby's reading score was 50 and his math score was 40, she still does not know how well Bobby is doing. In fact, she might get the false impression that he is better in reading when in fact 50 might be a very low score on the reading test and 40 may be a very good score on the math test. Now suppose Bobby's teacher also tells his mother that the average score (the mean, $\overline{X}$) on the reading test was 60, and the average score on the math test was 30. Aha! Now it looks as if Bobby is better in math than in reading. Further, if the standard deviation (*SD*) on both tests was 10, Bobby's true status becomes even more evident. Since his score in reading is exactly 1 *SD* below the mean (60 − 10 = 50), his *z* score is −1.00. On the other hand, his score in math is 1 *SD* above the mean (30 + 10 = 40) and his *z* score is +1.00. As shown, *z* scores can be translated into percentiles to show that Bobby is clearly better in math than in reading.[2]

	Raw Score	$\overline{X}$	*SD*	*z*	Percentile
Reading	50	60	10	−1.00	16th
Math	40	30	10	+1.00	84th

We can use Figure 14.2 to estimate percentile equivalents for given *z* scores, but this becomes more difficult for *z* scores that fall between the values given in the figure, for example, *z* scores of .63 or −1.78. A better approach is to use Table A.3 in Appendix A. For each *z* between −3.00 and +3.00 in the column labeled Area, Table A.3 gives the proportion of cases that are included up to that point. In other words, for any value of *z*, the area created to the left of the line on the curve represents the proportion of cases that falls below that *z* score. Thus, for *z* = .00 (the mean score), we go down the *z* columns until we come to .00 and we see that the corresponding area to the left is .5000, representing 50% of the cases and the 50th percentile. For Bobby, we simply go down the *z* columns until we come to −1.00 (his *z* score for reading), and we see that the corresponding area under the curve is .1587. By multiplying by 100 and rounding, we see that Bobby's reading score corresponds to approximately the 16th percentile (16% of the cases fall below *z* = −1.00). Similarly, for his math score of +1.00, the area is .8413, or approximately the 84th percentile.

[2]This analysis is based on the assumption that the same groups took the test and that the tests were normally distributed.

Of course there are lots of other fun things we can do with Table A.3. If we want to know the proportion of cases in the area to the right of a given z score, for example, we simply subtract the left area value from 1.00, since the total area under the curve equals all, or 1.00 (100% of the cases). So, if we want to know the percentage of students who did better than Bobby on the reading test, we subtract .1587 from 1.00 and we get .8413, or approximately 84%. Similarly, if we want to know the percentage of scores for any test that falls between $z = -1.00$ and $z = +1.00$, we subtract .1587 from .8413 and we get .6826, or approximately 68%. In other words, approximately 68% (68.26% to be exact) of the scores fall between $z = -1.00$ and $z = +1.00$ (as indicated by Figure 14.2, 34.13% + 34.13% = 68.26%). Thus, we can find the percentage of cases that falls between any two z scores by subtracting their Table A.3 area values.

Also by subtraction, we can find the percentage of cases that falls between the mean ($z = .00$, area = .5000) and any other z score. And, we can also reverse the process to find, for example, the z score that corresponds to a given percentile. To become a member of Mensa, for example, you have to have an IQ at or higher than the 98th percentile. The closest area value in Table A.3 that reflects 98th percentile is .9803, which corresponds to $z = 2.06$. In other words, your IQ has to be approximately two standard deviations ($+2\ \sigma$) above average, which corresponds to an IQ of 130 (see Figure 14.2).

Of course, as mentioned previously and as Table A.3 indicates, scores are not always exactly 1 *SD* (or 2 *SD* or 3 *SD*) above or below the mean. Usually we have to apply the following formula to convert a raw score to a z score:

$$Z = \frac{X - \overline{X}}{SD}, \text{ where } X \text{ is the raw score}$$

The only problem with z scores is that they involve negative numbers and decimals. It would be pretty hard to explain to Mrs. Bonker that her son was a −1.00. How do you tell a mother her son is a negative?! A simple solution is to transform z scores into T (or Z) scores. As Figure 14.2 indicates, z scores are actually the building blocks for a number of standard scores. Other standard scores represent transformations of z scores that communicate the same information in a more generally understandable form by eliminating negatives and/or decimals.

***T* Scores.** A ***T* score** (also called a **Z score**) is nothing more than a z score expressed in a different form. To transform a z score to a T score, you simply multiply the z score by 10 and add 50. In other words, $T = 10z + 50$. Thus, a z score of 0 (the mean score) becomes a T score of 50 [$T = 10(0) + 50 = 0 + 50 = 50$]. A z score of +1.00 becomes a T score of 60 [$T = 10(1.00) + 50 = 10 + 50 = 60$], and a z score of −1.00 becomes a T score of 40 [$T = 10(-1.00) + 50 = -10 + 50 = 40$]. Thus, when scores are transformed to T scores, the new distribution has a mean of 50 and a standard deviation of 10 (see Figure 14.2). It would clearly be much easier to communicate to Mrs. Bonker that Bobby is a 40 in reading and a 60 in math and that the average score is 50 than to tell her that he is a +1.00 and a −1.00 and the average score is .00.

If the raw score distribution is normal, then so is the z score distribution and the T score distribution. If, on the other hand, the original distribution is not normal (such as when a small sample group is involved), then neither are the z and T score distributions. In such cases the distribution resulting from the $10z + 50$ transformation is more accurately referred to as a Z distribution. However, even with a set of raw scores that are not normally distributed, we can produce a set of normalized Z scores. In either case, we can use the normal curve equivalencies to convert such scores into corresponding percentiles, or vice versa. As Figure 14.2 indicates, for example, a T of 50 = a percentile of 50%. Similarly, a T of 30 corresponds to a percentile of 2 and a T of 60 corresponds to the 84th percentile. The same is true for the other standard score transformations illustrated in Figure 14.2. The CEEB distribution is formed by multiplying T scores by 10 to eliminate decimals; it is calculated directly using CEEB = $100z + 500$. The AGCT (Army General Classification Test) distribution is formed by multiplying T scores by 2, and is formed directly using AGCT = $20z + 100$. In both cases, given values can be converted to percentiles (and vice versa) using normal curve equivalencies. Thus, a CEEB score of 400 corresponds to the 16th percentile and the 98th percentile corresponds to an AGCT score of 140.

Stanines. *Stanines* are standard scores that divide a distribution into nine parts. Stanine (short for "standard nine") equivalencies are derived using the formula $2z + 5$ and rounding resulting values to the nearest whole number. Stanines 2 through 8 each represent $\frac{1}{2}$ *SD* of the distribution; stanines 1 and 9 include the remainder. In other words, stanine 5 includes $\frac{1}{2}$ *SD* around the mean ($\overline{X}$); that is, it equals $\overline{X} \pm \frac{1}{4}$ *SD*. Stanine 6 goes from $+\frac{1}{4}$ *SD* to $+\frac{3}{4}$ *SD* ($\frac{1}{4}$ *SD* + $\frac{1}{2}$ *SD* = $\frac{3}{4}$ *SD*), and so forth. Stanine 1 includes any score that is less than $-1\frac{3}{4}$ *SD* (-1.75 *SD*) below the mean, and stanine 9 includes any score that is greater than $+1\frac{3}{4}$ *SD* ($+1.75$ *SD*) above the mean. As Figure 14.2 indicates (see the row of figures directly beneath the stanines), stanine 5 includes 20% of the scores, stanines 4 and 6 each contain 17%, stanines 3 and 7 each contain 12%, 2 and 8 each contain 7%, and 1 and 9 each contain 4% of the scores (percentages approximate). Thus, if a student was at the 7th stanine her percentile would be approximately $4 + 7 + 12 + 17 + 20 + 17 + 12 = 89$th percentile.

Like percentiles, stanines are very frequently reported in norms tables for standardized tests. They are very popular with school systems because they are so easy to understand and to explain to others. They are not as exact as other standard scores, but are useful for a variety of purposes. They are frequently used as a basis for grouping and are also used as a criterion for selecting students for special programs. A remediation program, for example, may select students who scored in the first and second stanine on a standardized reading test.

Use Figure 14.2 and Appendix A, Table A.3 to answer the following questions:

1. What percentile corresponds to a z score of $+2.00$?
2. What z score corresponds to a percentile of $-.20$?
3. Approximately what percentile corresponds to a stanine of 3?
4. What range of z scores encompasses 95% of the area in a normal curve?
5. What is the relationship of the mean, median, and mode in the normal distribution?

Measures of Relationship

Correlational research, the examination of the relationships between variables, was discussed in detail in Chapter 11. You will recall that correlational research involves collecting data to determine whether and to what degree a relationship exists between two or more quantifiable variables—not a causal relationship, just a relationship. Degree of relationship is expressed as a correlation coefficient, which is computed using two sets of scores from a single group of participants. The correlation coefficient provides an estimate of just how related two variables are. If two variables are highly related, a correlation coefficient near $+1.00$ (or -1.00) will be obtained; if two variables are not related, a coefficient near .00 will be obtained. There are a number of different methods of computing a correlation coefficient; which one is appropriate depends on the scale of measurement represented by the data. The two most frequently used correlational analyses are the rank difference correlation coefficient, usually referred to as the Spearman rho, and the product moment correlation coefficient, usually referred to as the Pearson *r*. See Table 11.2 for other correlational approaches, including the phi coefficient, biserial, and intraclass correlations, among others.

The Spearman Rho

The Spearman rho coefficient is used to correlate ranked data. The Spearman rho is thus appropriate when the data represent an ordinal scale (although it may be used with interval data) and is used when the median and quartile deviation are used. If only one of the variables to be correlated is ranked, the other variable to be correlated must also be expressed in terms of ranks. Thus, if intelligence were to be correlated with class rank, students' intelligence scores would have to be translated into ranks. If more than one participant receives the same score, then the corresponding ranks are averaged. So, for example, if two participants have the same highest score, they are each assigned the average of rank 1 and rank 2, namely, rank 1.5, and the next highest score is assigned rank 3. Similarly, the 24th and 25th highest scores, if identical, would each be assigned the rank 24.5. Like most other correlation coefficients, the Spearman rho produces a coefficient somewhere between −1.00 and +1.00. If, for example, a group of participants achieves identical ranks on both variables the coefficient will be +1.00.

The Pearson *r*

The Pearson *r* correlation coefficient is the most appropriate measure when the variables to be correlated are either interval or ratio. Like the mean and the standard deviation, the Pearson *r* takes into account each and every score in both distributions; it is also the most stable measure of correlation. Since most educational measures represent interval scales, the Pearson *r* is usually the most used coefficient for determining relationship. An assumption associated with the application of the Pearson *r* is that the relationship between the variables being correlated is a linear one. If this is not the case, the Pearson *r* will not yield a valid indication of relationship. If there is any question concerning the linearity of the relationship, the two sets of data should be plotted as previously shown in Figure 11.1.

CALCULATION FOR INTERVAL DATA

Because most educational data are represented in interval scales, we will calculate the measure of central tendency, variability, relationship, and relative position appropriate for interval data. There are several alternate formulas available for computing each of these measures; in each case, however, we will use the easiest, raw score formula. At first glance some of the formulas may look scary but they are really easy. The only reason they look hard is because they involve symbols with which you may be unfamiliar. As promised, however, each formula transforms "magically" into an arithmetic problem; all you have to do is substitute the correct numbers for the correct symbols.

Thus far most of this chapter has introduced varied types of statistics such as means, central tendency, variables, normal curves, and measures of relationship. These and other statistics are used to analyze various types of quantitative data. The remainder of this chapter and Chapter 15 will focus on quantitative data analysis. We will demonstrate aspects of data analysis in two ways. The first will be based on data analysis by following step-by-step procedures; the second will entail analysis by computer.

Although computer-based data analysis can be more efficient, it is important to work through data analyses for yourself to obtain a basic understanding of the research results. Once you are familiar with the step-by-step examples, you should also carry out analyses by computer.

For our computer analysis, we will use the SPSS student version 10.0 for Windows. SPSS is the most commonly used quantitative desktop computer analysis application. Please bear in mind that this is not a statistics course; our purpose is not to teach you how to use SPSS, but rather to illustrate how to use it to perform your quantitative data analysis. We will not examine all the analyses available in SPSS, but we will look at many of the basic ones commonly used in quantitative research. You will be able to compare the two types of analysis and see how they each can be used to produce the same results. (Note that the step-by-step and SPSS analyses do produce slightly different results due to the fact that the step-by-step analyses are worked out to two decimal places and the SPSS analyses are worked out variously to up to five decimal places. These differences do not significantly affect the results.)

Symbols

Before we start calculating let's get acquainted with a few basic statistical symbols. First, X (without a bar) is usually used to symbolize a score. If you see a column of numbers, and at the top of that column is an X, you know that the column represents a set of scores. If there are two sets of scores they may be labeled X_1 and X_2 or X and Y, it does not matter which.

Another symbol used frequently is the Greek letter Σ, which is used to indicate addition. Σ means "the sum of," or "add them all up." Thus ΣX means "add up all the Xs" and ΣY means "add up all the Ys." Isn't this easy? Now, if any symbol has a bar over it, such as $\overline{X}$, that indicates the mean, or arithmetic average, of the scores. Thus $\overline{X}$ refers to the mean of the X scores and $\overline{Y}$ refers to the mean of the Y scores.

A capital N refers to the number of participants; $N = 20$ means that there are 20 participants (N is for number; makes sense, doesn't it?). If one analysis involves several groups, the number of participants in each group is indicated with a lowercase letter n and a subscript indicating the group. If there are three groups, and the first group has 15 participants, the second group has 18, and the third group has 20, this is symbolized as $n_1 = 15$, $n_2 = 18$, and $n_3 = 20$. The total number of subjects is represented with a capital $N = 53$ ($15 + 18 + 20 = 53$).

Finally, you must get straight the difference between ΣX^2 and $(\Sigma X)^2$; they do not mean the same thing. Different formulas may include one or the other or both and it is very important to interpret each correctly, since a formula tells you what to do and you must do exactly what it tells you. Now let us look at ΣX^2. What does it tell you? The Σ tells you that you are supposed to add something up. What you are supposed to add up are X^2s. What do you suppose X^2 means? Right. It means the square of the score; if $X = 4$, then $X^2 = 4 \times 4 = 16$. Thus, ΣX^2 says, square each score and then add up all the squares. Now let us look at $(\Sigma X)^2$. Since whatever is in the parentheses is always done first, the first thing we do is ΣX. You already know what that means: "add up all the scores." And then what? Right. You add up all the scores and then you square the total. As an example:

X	X^2	
1	1	
2	4	$\Sigma X^2 = 55$
3	9	
4	16	$(\Sigma X)^2 = 225$
5	25	
$\Sigma X = 15$	$\Sigma X^2 = 55$	

As you can see, there is a big difference between ΣX^2 and $(\Sigma X)^2$, so watch out! To summarize, symbols commonly used in statistical formulas are as follows:

X = a score
Σ = the sum of; add them up
ΣX = the sum of all the scores
$\overline{X}$ = the mean, or arithmetic average, of the scores
N = total number of subjects
n = number of subjects in a particular group
ΣX^2 = the sum of the squares; square each score and add up all the squares
$(\Sigma X)^2$ = the square of the sum; add up the scores and square the sum, or total

If you approach each statistic in an orderly fashion, it makes your statistical life easier. A suggested procedure is as follows:

1. Make the columns required by the formula (e.g., X, X^2, as just shown) and find the sum of each column.
2. Label the sum of each column; in the previous example, the label for the sum of the X column = ΣX, and the label for the sum of the X^2 column = ΣX^2.
3. Write the formula.
4. Write the arithmetic equivalent of the formula (e.g., $(\Sigma X)^2 = (15)^2$).
5. Solve the arithmetic problem (e.g., $(15)^2 = 225$).

The Mean

Although sample sizes of 5 are hardly ever considered to be acceptable, we will use this number of participants for illustration purposes. Our calculations will be based on the scores of 5 participants so that you can concentrate on how the calculation is being done and will not get lost in the numbers. For the same reason we will also use small numbers. Now, assume we have the following scores for some old friends of ours and we want to compute the mean, or arithmetic average.

	X
Iggie	1
Hermie	2
Fifi	3
Teenie	4
Tiny	5

Remember that a column labeled X means "here come the scores!"

The formula for the mean is $\overline{X} = \dfrac{\Sigma X}{N}$

You are now looking at a statistic. Looks bad, right? Now let us first see what it really says. It reads "the mean ($\overline{X}$) is equal to the sum of the scores (ΣX) divided by the number of participants (N)." So, to find X we need ΣX and N.

X
1
2
3
4
5
$\Sigma X = 15$

Clearly, $\Sigma X = 1 + 2 + 3 + 4 + 5 = 15$
$N = 5$ (there are 5 participants, right?)

Now we have everything we need to find the mean and all we have to do is substitute the correct number for each symbol.

$$\overline{X} = \frac{\Sigma X}{N} = \frac{15}{5} = 3$$

Now what do we have? Right! An arithmetic problem. And hardly a difficult one! More like an elementary school arithmetic problem. And all we did was to substitute each symbol with the appropriate number. Thus,

$$\overline{X} = \frac{\Sigma X}{N} = \frac{15}{5} = 3(3.00)$$

and the mean is equal to 3.00. If you look at the scores you can see that 3 is clearly the average score. Since traditionally statistical results are given with 2 decimal places, our "official" answer is 3.00.

Was that hard? Cer-tain-ly not! And guess what—you just learned how to do a statistic! Are they all going to be that easy? Of course!

The Standard Deviation

Earlier we explained the fact that the standard deviation is the square root of the variance, which is based on the distance of each score from the mean. To calculate the standard deviation (*SD*), however, we do not have to calculate variance scores; we can use a raw score formula that gives us the same answer with less grief. Now, before you look at the formula, remember that no matter how bad it looks, it is going to turn into an easy arithmetic problem. Ready?

$$SD = \sqrt{\frac{SS}{N-1}} \text{ where } SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

or

$$SD = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}}$$

In other words, the *SD* is equal to the square root of the sum of squares (*SS*) divided by $N - 1$.

If the standard deviation of a population is being calculated, the formula is exactly the same, except we divide sum of squares by *N*, instead of $N - 1$. The reason is that a sample standard deviation is considered to be a biased estimate of the population standard deviation. When we select a sample, especially a small sample, the probability is that participants will come from the middle of the distribution and that extreme scores will not be represented. Thus, the range of sample scores will be smaller than the population range, as will be the sample standard deviation. As the sample size increases, so do the chances of getting extreme scores; thus, the smaller the sample, the more important it is to correct for the downward bias. By dividing by $N - 1$ instead of *N*, we make the denominator (bottom part!) smaller, and thus $\frac{SS}{N-1}$ is larger, closer to the population *SD* than $\frac{SS}{N}$. For example, if $SS = 18$ and $N = 10$, then

$$\frac{SS}{N-1} = \frac{18}{9} = 2.00 \qquad \frac{SS}{N} = \frac{18}{8} = 1.80$$

Now just relax and look at each piece of the formula; you already know what each one means. Starting with the easy one, *N* refers to what? Right—the number of subjects. How about (ΣX)? Right—the sum of the scores. And $(\Sigma X)^2$? Right—the square of the sum of the scores. That leaves ΣX^2, which means the sum of what? . . . Fantastic. The sum of the squares. Okay, let us use the same scores we used to calculate the mean. The first thing we need to do is to square each score and then add those squares up—while we are at it we can also go ahead and add up all the scores.

	X	X^2	
Iggie	1	1	
Hermie	2	4	$\Sigma X = 15$
Fifi	3	9	$\Sigma X^2 = 55$
Teenie	4	16	$N = 5$
Tiny	5	25	$N - 1 = 4$
	$\Sigma X = 15$	$\Sigma X^2 = 55$	

Do we have everything we need? Yes. Does the formula ask for anything else? No. We are in business. Substituting each symbol with its numerical equivalent we get

$$SS = \frac{(\Sigma X)^2}{N} = 55 - \frac{(15)^2}{5}$$

Now what do we have? A statistic? No! An arithmetic problem? Yes! A hard arithmetic problem? No! It is harder than 15/5 but it is not hard. If we just do what the formula tells us to do we will have no problem at all. The first thing it tells us to do is to square 15:

$$SS = \Sigma X^2 \frac{(\Sigma X)^2}{N} = 55 - \frac{(15)^2}{5} = 55 - \frac{225}{5}$$

So far so good. The next thing the formula tells us to do is divide 225 by 5, which equals 45. It is looking a lot better; now it is really an easy arithmetic problem. Okay, the next step is to subtract 45 from 55 and we get a sum of squares (*SS*) equal to 10.00.

Mere child's play. Think you can figure out the next step? Terrific! Now that we have *SS*, we simply substitute it into the *SD* formula as follows:

$$SD = \sqrt{\frac{SS}{N-1}} = \sqrt{\frac{10}{4}} = \sqrt{2.5}$$

To find the square root of 2.5, simply enter 2.5 into your calculator and hit the square root button ($\sqrt{}$); the square root of 2.5 is 1.58. Substituting in our square root we have

$$SD = \sqrt{2.5} = 1.58$$

and the standard deviation is 1.58. If we had calculated the standard deviation for the IQ distribution shown in Figure 14.2, what would we have gotten? Right, 15. Now you know how to do two useful descriptive statistics.

Standard Scores

Your brain has earned a rest, and the formula for a *z* score is a piece of cake:

$$z = \frac{X - \overline{X}}{SD}$$

To convert scores to *z* scores we simply apply that formula to each score. We have already computed the mean and the standard deviation for the following scores:

	X	
Iggie	1	
Hermie	2	$X = 3$
Fifi	3	
Teenie	4	$SD = 1.58$
Tiny	5	

Let's see how Iggie's *z* score works out:

$$\text{Iggie} \quad z = \frac{X - \overline{X}}{SD} = \frac{1-3}{1.58} = \frac{-2}{1.58} = -1.26$$

This tells us that Iggie's standard score was 1.26 standard deviations below average. In case you have forgotten, if the signs are the same (two positives or two negatives), the answer in a multiplication or division problem is a positive number; if the signs are different, the answer is a negative number, as in Iggie's case. For the rest of our friends, the results are

$$\text{Hermie} \quad z = \frac{X - \overline{X}}{SD} = \frac{-1}{1.58} = -.63$$

$$\text{Fifi} \quad z = \frac{X - \overline{X}}{SD} = \frac{0}{1.58} = 0.0$$

$$\text{Teenie} \quad z = \frac{X - \overline{X}}{SD} = \frac{1}{1.58} = +.63$$

$$\text{Tiny} \quad z = \frac{X - \overline{X}}{SD} = \frac{2}{1.58} = +1.26$$

Notice that Fifi's score was the same as the mean score. Her z score is .00, meaning that her score is no distance from the mean (it's on the mean). On the other hand, Iggie's and Hermie's scores were below the mean so their z scores are negative, whereas Teenie and Tiny scored above the mean and their z scores are positive. If we want to eliminate the negatives, we can convert each z score to a Z score. Remember those? Multiplying each z score by 10 and adding 50 gives $Z = 10z + 50$. If we apply the z score formula we get

$$\begin{aligned}
\text{Iggie} \quad Z &= 10z + 50 = 10(-1.26) + 50 \\
&= -12.6 + 50 \\
&= 50 - 12.6 \\
&= 37.40 \\
\text{Hermie} \quad Z &= 10z + 50 = 10(-.63) + 50 \\
&= -6.3 + 50 \\
&= 50 - 6.3 \\
&= 43.70 \\
\text{Fifi} \quad Z &= 10z + 50 = 10(.00) + 50 \\
&= .00 + 50 \\
&= 50 + .00 \\
&= 50.00 \\
\text{Teenie} \quad Z &= 10z + 50 = 10(+.63) + 50 \\
&= 6.3 + 50 \\
&= 50 + 6.3 \\
&= 56.30 \\
\text{Tiny} \quad Z &= 10z + 50 = 10(+1.26) + 50 \\
&= 12.6 + 50 \\
&= 50 + 12.6 \\
&= 62.60
\end{aligned}$$

Note the analyses produced by these three common statistics: the mean = 3; the standard deviation = 1.58; and standard scores are −1.26, −.63, .00; .63., and 1.26. Now we will illustrate the use of SPSS to obtain the same results from the previous analysis. In this chapter and the next, we will show you both the analyses generated by the step-by-step approach and the SPSS approach.

Obtaining Descriptive Statistics with SPSS 10.0

When the data set is large, it is often easier to input, or type, the scores into a spreadsheet and generate the statistics you want using a computer program such as the Statistical Package for the Social Sciences (SPSS.) Figure 14.4 shows the same five scores for our friends as we just discussed. This time they have been entered into the SPSS spreadsheet.

To generate the descriptive statistics you want click on the "Analyze" menu and choose the "Descriptive Statistics" option as shown in Figure 14.5. In the Descriptive Statistics menu choose the "Descriptives . . ." option.