Tools You Will Need

The following items are considered essential background material for this chapter. If you doubt your knowledge of any of these items, you should review the appropriate chapter or section before proceeding.

- Proportions (math review, Appendix A)
- Fractions
- Decimals
- Percentages

- Scales of measurement (Chapter 1):
  Nominal, ordinal, interval, and ratio

- Continuous and discrete variables (Chapter 1)
- Real limits (Chapter 1)


2.1 INTRODUCTION TO FREQUENCY DISTRIBUTIONS

When the data collection phase of a research study is completed, the results usually consist of pages of numbers. The immediate problem for the researcher is to organize the scores into some comprehensible form so that any patterns in the data can be seen easily and communicated to others. This is the job of descriptive statistics: to simplify the organization and presentation of data. One of the most common procedures for organizing a set of data is to place the scores in a frequency distribution.

**DEFINITION**

A **frequency distribution** is an organized tabulation of the number of individuals located in each category on the scale of measurement.

A frequency distribution takes a disorganized set of scores and places them in order from highest to lowest, grouping together all individuals who have the same score. If the highest score is $X = 10$, for example, the frequency distribution groups together all the 10s, then all the 9s, then the 8s, and so on. Thus, a frequency distribution allows the researcher to see "at a glance" the entire set of scores. It shows whether the scores are generally high or low, whether they are concentrated in one area or spread out across the entire scale, and generally provides an organized picture of the data. In addition to providing a picture of the entire set of scores, a frequency distribution allows you to see the location of any individual score relative to all of the other scores in the set.

A frequency distribution can be structured either as a table or as a graph, but in either case the distribution presents the same two elements:

1. The set of categories that make up the original measurement scale.
2. A record of the frequency, or number of individuals in each category.

Thus, a frequency distribution presents a picture of how the individual scores are distributed on the measurement scale—hence the name frequency distribution.

2.2 FREQUENCY DISTRIBUTION TABLES

It is customary to list categories from highest to lowest, but this is an arbitrary arrangement. Many computer programs list categories from lowest to highest.

The simplest frequency distribution table presents the measurement scale by listing the different measurement categories ($X$ values) in a column from highest to lowest. Beside each $X$ value, we indicate the frequency, or the number of times that particular measurement occurred in the data. It is customary to use an $X$ as the column heading for the scores and an $f$ as the column heading for the frequencies. An example of a frequency distribution table follows.

**EXAMPLE 2.1**

The following set of $N = 20$ scores was obtained from a 10-point statistics quiz. We will organize these scores by constructing a frequency distribution table. Scores:

8, 9, 8, 7, 10, 9, 6, 4, 9, 8,
7, 8, 10, 9, 8, 6, 9, 7, 8, 8
The highest score is \( X = 10 \), and the lowest score is \( X = 4 \). Therefore, the first column of the table lists the categories that make up the scale of measurement (\( X \) values) from 10 down to 4. Notice that all of the possible values are listed in the table. For example, no one had a score of \( X = 5 \), but this value is included. With an ordinal, interval, or ratio scale, the categories are listed in order (usually highest to lowest). For a nominal scale, the categories can be listed in any order.

2. The frequency associated with each score is recorded in the second column. For example, two people had scores of \( X = 6 \), so there is a 2 in the \( f \) column beside \( X = 6 \).

Because the table organizes the scores, it is possible to see very quickly the general quiz results. For example, there were only two perfect scores, but most of the class had high grades (8s and 9s). With one exception (the score of \( X = 4 \)), it appears that the class has learned the material fairly well.

Notice that the \( X \) values in a frequency distribution table represent the scale of measurement, not the actual set of scores. For example, the \( X \) column lists the value 10 only one time, but the frequency column indicates that there are actually two values of \( X = 10 \). Also, the \( X \) column lists a value of \( X = 5 \), but the frequency column indicates that no one actually had a score of \( X = 5 \).

You also should notice that the frequencies can be used to find the total number of scores in the distribution. By adding up the frequencies, you obtain the total number of individuals:

\[
\sum f = N
\]

There may be times when you need to compute the sum of the scores, \( \sum X \), or perform other computations for a set of scores that has been organized into a frequency distribution table. To complete these calculations correctly, you must use the frequencies presented in the table. That is, it is essential to use the information in the \( f \) column as well as the \( X \) column to obtain the full set of scores.

When it is necessary to perform calculations for scores that have been organized into a frequency distribution table, the safest procedure is to take the individual scores out of the table before you begin any computations. This process is demonstrated in the following example.

\[\text{Example 2.2}\]

Consider the frequency distribution table shown in the margin. The table shows that the distribution has one 5, two 4s, three 3s, three 2s, and one 1. If you simply list all the individual scores, you can safely proceed with calculations such as finding \( \Sigma X \) or \( \Sigma X^2 \). Note that the complete set contains \( N = \Sigma f = 10 \) scores and your list should contain 10 values. For example, to compute \( \Sigma X \) you must add all 10 scores:

\[
\Sigma X = 5 + 4 + 4 + 3 + 3 + 3 + 2 + 2 + 2 + 1
\]

For the distribution in this table, you should obtain \( \Sigma X = 29 \). Try it yourself. Similarly, to compute \( \Sigma X^2 \) you square each of the 10 scores and then add the squared values.
\[ \Sigma X^2 = 5^2 + 4^2 + 4^2 + 3^2 + 3^2 + 3^2 + 2^2 + 2^2 + 1^2 \]

This time you should obtain \( \Sigma X^2 = 97 \).

An alternative way to get \( \Sigma X \) from a frequency distribution table is to multiply each \( X \) value by its frequency and then add these products. This sum may be expressed in symbols as \( \Sigma fX \). The computation is summarized as follows for the data in Example 2.2:

<table>
<thead>
<tr>
<th>( X )</th>
<th>( f )</th>
<th>( fX )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>5 (the one 5 totals 5)</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>8 (the two 4s total 8)</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>9 (the three 3s total 9)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6 (the three 2s total 6)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1 (the one 1 totals 1)</td>
</tr>
</tbody>
</table>

\[ \Sigma X = 29 \]

No matter which method you use to find \( \Sigma X \), the important point is that you must use the information given in the frequency column in addition to the information in the \( X \) column.

**PROPORTIONS AND PERCENTAGES**

In addition to the two basic columns of a frequency distribution, there are other measures that describe the distribution of scores and can be incorporated into the table. The two most common are proportion and percentage.

Proportion measures the fraction of the total group that is associated with each score. In Example 2.2, there were two individuals with \( X = 4 \). Thus, 2 out of 10 people had \( X = 4 \), so the proportion would be \( \frac{2}{10} = 0.20 \). In general, the proportion associated with each score is

\[ \text{proportion} = p = \frac{f}{N} \]

Because proportions describe the frequency \( (f) \) in relation to the total number \( (N) \), they often are called *relative frequencies*. Although proportions can be expressed as fractions (for example, \( \frac{2}{10} \)), they more commonly appear as decimals. A column of proportions, headed with a \( p \), can be added to the basic frequency distribution table (see Example 2.3).

In addition to using frequencies \( (f) \) and proportions \( (p) \), researchers often describe a distribution of scores with percentages. For example, an instructor might describe the results of an exam by saying that 15\% of the class earned A's, 23\% B's, and so on. To compute the percentage associated with each score, you first find the proportion \( (p) \) and then multiply by 100:

\[ \text{percentage} = p(100) = \frac{f}{N}(100) \]

Percentages can be included in a frequency distribution table by adding a column headed with \% (see Example 2.3).
EXAMPLE 2.3

The frequency distribution table from Example 2.2 is repeated here. This time we have added columns showing the proportion \( p \) and the percentage \( % \) associated with each score.

<table>
<thead>
<tr>
<th>( X )</th>
<th>( f )</th>
<th>( p = \frac{f}{N} )</th>
<th>( % = p \times 100 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>1/10 = 0.10</td>
<td>10%</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2/10 = 0.20</td>
<td>20%</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3/10 = 0.30</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3/10 = 0.30</td>
<td>30%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1/10 = 0.10</td>
<td>10%</td>
</tr>
</tbody>
</table>

LEARNING CHECK

1. Construct a frequency distribution table for the following set of scores.

Scores: 3, 4, 3, 2, 5, 3, 2, 1, 2, 5, 3, 4, 4, 2, 3

2. Find each of the following values for the sample in the following frequency distribution table.

<table>
<thead>
<tr>
<th>( \Sigma X )</th>
<th>( \Sigma X^2 )</th>
</tr>
</thead>
</table>
a. \( n \)     | 5               |
b. \( \Sigma X \) | 4               |
c. \( \Sigma X^2 \) | 3, 4           |
   | 2, 3           |
   | 1, 1           |

ANSWERS

1. | \( \Sigma X \) | \( f \) |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

2. a. \( n = 11 \)  
b. \( \Sigma X = 32 \)  
c. \( \Sigma X^2 = 106 \) (square then add all 11 scores)

GROUPED FREQUENCY DISTRIBUTION TABLES

When a set of data covers a wide range of values, it is unreasonable to list all the individual scores in a frequency distribution table. For example, a set of exam scores ranges from a low of \( X = 41 \) to a high of \( X = 96 \). These scores cover a range of more than 50 points.

If we were to list all the individual scores from \( X = 96 \) down to \( X = 41 \), it would take 56 rows to complete the frequency distribution table. Although this would organize the data, the table would be long and cumbersome. Remember: The purpose for constructing a table is to obtain a relatively simple, organized picture of the data. This can be accomplished by grouping the scores into intervals and then listing the
When the scores are whole numbers, the total number of rows for a regular table can be obtained by finding the difference between the highest and the lowest scores and adding 1:

\[
\text{rows} = \text{highest} - \text{lowest} + 1
\]

**RULE 1**

The grouped frequency distribution table should have about 10 class intervals. If a table has many more than 10 intervals, it becomes cumbersome and defeats the purpose of a frequency distribution table. On the other hand, if you have too few intervals, you begin to lose information about the distribution of the scores. At the extreme, with only one interval, the table would not tell you anything about how the scores are distributed. Remember that the purpose of a frequency distribution is to help a researcher see the data. With too few or too many-intervals, the table will not provide a clear picture. You should note that 10 intervals is a general guide. If you are constructing a table on a blackboard, for example, you probably want only 5 or 6 intervals. If the table is to be printed in a scientific report, you may want 12 or 15 intervals. In each case, your goal is to present a table that is relatively easy to see and understand.

**RULE 2**

The width of each interval should be a relatively simple number. For example, 2, 5, 10, or 20 would be a good choice for the interval width. Notice that it is easy to count by 5s or 10s. These numbers are easy to understand and make it possible for someone to see quickly how you have divided the range.

**RULE 3**

The bottom score in each class-interval should be a multiple of the width. If you are using a width of 10 points, for example, the intervals should start with 10, 20, 30, 40, and so on. Again, this makes it easier for someone to understand how the table has been constructed.

**RULE 4**

All intervals should be the same width. They should cover the range of scores completely with no gaps and no overlaps, so that any particular score belongs in exactly one interval.

The application of these rules is demonstrated in Example 2.4.

**EXAMPLE 2.4**

An instructor has obtained the set of \( N = 25 \) exam scores shown here. To help organize these scores, we will place them in a frequency distribution table. The scores are:

\[
82, 75, 88, 93, 53, 84, 87, 58, 72, 94, 69, 84, 61, \\
91, 64, 87, 84, 70, 76, 89, 75, 80, 73, 78, 60
\]

The first step is to determine the range of scores. For these data, the smallest score is \( X = 53 \) and the largest score is \( X = 94 \), so a total of 42 rows would be needed for a table that lists each individual score. Because 42 rows would not provide a simple table, we have to group the scores into class intervals.

The best method for finding a good interval width is a systematic trial-and-error approach that uses rules 1 and 2 simultaneously. According to rule 1, we want about
Because the bottom interval usually extends below the lowest score and the top interval extends beyond the highest score, you often will need slightly more than the computed number of intervals.

10 intervals; according to rule 2, we want the interval width to be a simple number. For this example, the scores cover a range of 42 points, so we will try several different interval widths to see how many intervals are needed to cover this range. For example, if each interval is 2 points wide, it would take 21 intervals to cover a range of 42 points. This is too many. What about an interval width of 5? What about a width of 10? The following table shows how many intervals would be needed for these possible widths:

<table>
<thead>
<tr>
<th>Width</th>
<th>Number of Intervals Needed to Cover a Range of 42 Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>21 (too many)</td>
</tr>
<tr>
<td>5</td>
<td>9 (OK)</td>
</tr>
<tr>
<td>10</td>
<td>5 (too few)</td>
</tr>
</tbody>
</table>

Notice that an interval width of 5 will result in about 10 intervals, which is exactly what we want.

The next step is to actually identify the intervals. The lowest score for these data is $X = 53$, so the lowest interval should contain this value. Because the interval should have a multiple of 5 as its bottom score, the interval should begin at 50. The interval has a width of 5, so it should contain 5 values: 50, 51, 52, 53, and 54. Thus, the bottom interval is 50–54. The next interval would start at 55 and go to 59.

The complete frequency distribution table showing all of the class intervals is presented in Table 2.1.

Once the class intervals are listed, you complete the table by adding a column of frequencies. The values in the frequency column indicate the number of individuals whose scores are located in that class interval. For this example, there were three students with scores in the 60–64 interval, so the frequency for this class interval is $f = 3$ (see Table 2.1). The basic table can be extended by adding columns showing the proportion and percentage associated with each class interval.

Finally, you should note that after the scores have been placed in a grouped table, you lose information about the specific value for any individual score. For example, Table 2.1 shows that one person had a score between 65 and 69, but the table does not identify the exact value for the score. In general, the wider the class intervals are, the more information is lost. In Table 2.1 the interval width is 5 points, and the table

**TABLE 2.1**

A grouped frequency distribution table showing the data from Example 2.4.
The original scores range from a high of $X = 94$ to a low of $X = 53$. This range has been divided into 9 intervals with each interval exactly 5 points wide. The frequency column ($f$) lists the number of individuals with scores in each of the class intervals.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90–94</td>
<td>3</td>
</tr>
<tr>
<td>85–89</td>
<td>4</td>
</tr>
<tr>
<td>80–84</td>
<td>5</td>
</tr>
<tr>
<td>75–79</td>
<td>4</td>
</tr>
<tr>
<td>70–74</td>
<td>3</td>
</tr>
<tr>
<td>65–69</td>
<td>1</td>
</tr>
<tr>
<td>60–64</td>
<td>3</td>
</tr>
<tr>
<td>55–59</td>
<td>1</td>
</tr>
<tr>
<td>50–54</td>
<td>1</td>
</tr>
</tbody>
</table>
shows that there are three people with scores in the lower 60s and one person with a score in the upper 60s. This information would be lost if the interval width were increased to 10 points. With an interval width of 10, all of the 60s would be grouped together into one interval labeled 60–69. The table would show a frequency of four people in the 60–69 interval, but it would not tell whether the scores were in the upper 60s or the lower 60s.

Recall from Chapter 1 that a continuous variable has an infinite number of possible values and can be represented by a number line that is continuous and contains an infinite number of points. However, when a continuous variable is measured, the resulting measurements correspond to intervals on the number line rather than single points. If you are measuring time in seconds, for example, a score of $X = 8$ seconds actually represents an interval bounded by the real limits 7.5 seconds and 8.5 seconds. Thus, a frequency distribution table showing a frequency of $f = 3$ individuals all assigned a score of $X = 8$ does not mean that all three individuals had exactly the same measurement. Instead, you should realize that the three measurements are simply located in the same interval between 7.5 and 8.5.

The concept of real limits also applies to the class intervals of a grouped frequency distribution table. For example, a class interval of 40–49 contains scores from $X = 40$ to $X = 49$. These values are called the apparent limits of the interval because it appears that they form the upper and lower boundaries for the class interval. But $X = 40$ is actually an interval from 39.5 to 40.5. Similarly, $X = 49$ is an interval from 48.5 to 49.5. Therefore, the real limits of the interval are 39.5 (the lower real limit) and 49.5 (the upper real limit). Notice that the next higher class interval is 50–59, which has a lower real limit of 49.5. Thus, the two intervals meet at the real limit 49.5, so there are no gaps in the scale. You also should notice that the width of each class interval becomes easier to understand when you consider the real limits of an interval. For example, the interval 50–59 has real limits of 49.5 and 54.5. The distance between these two real limits (10 points) is the width of the interval.

### LEARNING CHECK

1. For each of the following situations, determine what interval width is most appropriate for a grouped frequency distribution and identify the apparent limits of the bottom interval.
   - a. Scores range from $X = 7$ to $X = 21$.
   - b. Scores range from $X = 52$ to $X = 98$.
   - c. Scores range from $X = 16$ to $X = 93$.

2. Using only the frequency distribution table presented in Table 2.1, how many individuals had a score of $X = 73$?

### ANSWERS

1. a. A width of 2 points would require 8 intervals. Bottom intervals is 6–7.
   - b. A width of 5 points would require 10 intervals. Bottom intervals is 50–54.
   - c. A width of 10 points would require 9 intervals. Bottom intervals is 10–19.

2. After a set of scores has been summarized in a grouped table, you cannot determine the frequency for any specific score. There is no way to determine how many individuals had $X = 73$ from the table alone. (You can say that at most three people had $X = 73$.)
2.3 FREQUENCY DISTRIBUTION GRAPHS

A frequency distribution graph is basically a picture of the information available in a frequency distribution table. We will consider several different types of graphs, but all start with two perpendicular lines called axes. The horizontal line is called the X-axis, or the abscissa (ab-SIS-uh). The vertical line is called the Y-axis, or the ordinate. The measurement scale (set of X values) is listed along the X-axis with values increasing from left to right. The frequencies are listed on the Y-axis with values increasing from bottom to top. As a general rule, the point where the two axes intersect should have a value of zero for both the scores and the frequencies. A final general rule is that the graph should be constructed so that its height (Y-axis) is approximately two-thirds to three-quarters of its length (X-axis). Violating these guidelines can result in graphs that give a misleading picture of the data (see Box 2.1).

When the data consist of numerical scores that have been measured on an interval or ratio scale, there are two options for constructing a frequency distribution graph. The two types of graph are called histograms and polygons.

Histograms To construct a histogram, you first list the numerical scores (the categories of measurement) along the X-axis. Then you draw a bar above each X value so that

a. The height of the bar corresponds to the frequency for that category.
b. The width of the bar extends to the real limits of the category.

Because the bars extend to the real limits for each category, adjacent bars touch so that there are no spaces or gaps between bars. An example of a histogram is shown in Figure 2.1.

When data have been grouped into class intervals, you can construct a frequency distribution histogram by drawing a bar above each interval so that the width of the bar extends to the real limits of the interval (the lower real limit of the lowest score and the upper real limit of the highest score in the interval). This process is demonstrated in Figure 2.2.

For the two histograms shown in Figures 2.1 and 2.2, notice that the values on both the vertical and horizontal axes are clearly marked and that both axes are labeled. Also note that, whenever possible, the units of measurement are specified; for example, Figure 2.2 shows a distribution of heights measured in inches. Finally, notice that the horizontal axis in Figure 2.2 does not list all of the possible heights starting from zero and going up to 48 inches. Instead, the graph clearly shows a break between zero and 30, indicating that some scores have been omitted.

A modified histogram A slight modification to the traditional histogram produces a very easy to draw and simple to understand sketch of a frequency distribution. Instead of drawing a bar above each score, the modification consists of drawing a stack of blocks. Each block represents one individual, so the number of blocks above each score corresponds to the frequency for that score. An example is shown in Figure 2.3.

Note that the number of blocks in each stack makes it very easy to see the absolute frequency for each category. In addition it is easy to see the exact difference in frequency from one category to another. In Figure 2.3, for example, it is easy to see that there are exactly two more people with scores of X = 2 than with scores of X = .1.
Because the frequencies are clearly displayed by the number of blocks, this type of display eliminates the need for a vertical line (the Y-axis) showing frequencies. In general, this kind of graph provides a simple and concrete picture of the distribution for a sample of scores. Note that we often will use this kind of graph to show sample data throughout the rest of the book. You should also note, however, that this kind of display simply provides a sketch of the distribution and is not a substitute for an accurately drawn histogram with two labeled axes.
Polygons The second option for graphing a distribution of numerical scores from an interval or ratio scale of measurement is called a polygon. To construct a polygon, you begin by listing the numerical scores (the categories of measurement) along the X-axis. Then,

a. A dot is centered above each score so that the vertical position of the dot corresponds to the frequency for the category.

b. A continuous line is drawn from dot to dot to connect the series of dots.

c. The graph is completed by drawing a line down to the X-axis (zero frequency) at each end of the range of scores. The final lines are usually drawn so that they reach the X-axis at a point that is one category below the lowest score on the left side and one category above the highest score on the right side. An example of a polygon is shown in Figure 2.4.

A polygon also can be used with data that have been grouped into class intervals. For a grouped distribution, you position each dot directly above the midpoint of the class interval. The midpoint can be found by averaging the highest and the lowest scores in the interval. For example, a class interval that is listed as 20–29 would have a midpoint of 24.5.

\[
\text{midpoint} = \frac{20 + 29}{2} = \frac{49}{2} = 24.5
\]

An example of a frequency distribution polygon with grouped data is shown in Figure 2.5.

---

**GRAPHS FOR NOMINAL OR ORDINAL DATA**

When the scores are measured on a nominal or ordinal scale (usually nonnumerical values), the frequency distribution can be displayed in a bar graph.

---

**FIGURE 2.4**

An example of a frequency distribution polygon. The same set of data is presented in a frequency distribution table and in a polygon.

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
FIGURE 2.5
An example of a frequency distribution polygon for grouped data. The same set of data is presented in a grouped frequency distribution table and in a polygon.

<table>
<thead>
<tr>
<th>Scores</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-5</td>
<td>2</td>
</tr>
<tr>
<td>6-7</td>
<td>3</td>
</tr>
<tr>
<td>8-9</td>
<td>3</td>
</tr>
<tr>
<td>10-11</td>
<td>5</td>
</tr>
<tr>
<td>12-13</td>
<td>4</td>
</tr>
</tbody>
</table>

Bar graphs  A bar graph is essentially the same as a histogram, except that spaces are left between adjacent bars. For a nominal scale, the space between bars emphasizes that the scale consists of separate, distinct categories. For ordinal scales, separate bars are used because you cannot assume that the categories are all the same size.

To construct a bar graph, list the categories of measurement along the X-axis and then draw a bar above each category so that the height of the bar corresponds to the frequency for the category. An example of a bar graph is shown in Figure 2.6.

FIGURE 2.6
A bar graph showing the distribution of personality types in a sample of college students. Because personality type is a discrete variable measured on a nominal scale, the graph is drawn with space between the bars.

When you can obtain an exact frequency for each score in a population, you can construct frequency distribution graphs that are exactly the same as the histograms, polygons, and bar graphs that are typically used for samples. For example, if a population is defined as a specific group of N = 50 people, we could easily determine how many have IQs of X = 110. However, if we are interested in the entire population of adults in the United States, it would be impossible to obtain an exact count of the number of people with an IQ of 110. Although it is still possible to construct graphs showing frequency distributions for extremely large populations, the graphs usually involve two special features: relative frequencies and smooth curves.
Relative frequencies Although you usually cannot find the absolute frequency for each score in a population, you very often can obtain relative frequencies. For example, you may not know exactly how many fish are in the lake, but after years of fishing you do know that there are twice as many bluegill as there are bass. You can represent these relative frequencies in a bar graph by making the bar above bluegill two times taller than the bar above bass (Figure 2.7). Notice that the graph does not show the absolute number of fish. Instead, it shows the relative number of bluegill and bass.

Smooth curves When a population consists of numerical scores from an interval or a ratio scale, it is customary to draw the distribution with a smooth curve instead of the jagged, step-wise shapes that occur with histograms and polygons. The smooth curve indicates that you are not connecting a series of dots (real frequencies) but instead are showing the relative changes that occur from one score to the next. One commonly occurring population distribution is the normal curve. The word normal refers to a specific shape that can be precisely defined by an equation. Less precisely, we can describe a normal distribution as being symmetrical, with the greatest frequency in the middle and relatively smaller frequencies as you move toward either extreme. A good example of a normal distribution is the population distribution for IQ scores shown in Figure 2.8. Because normal-shaped distributions occur commonly and because this
shape is mathematically guaranteed in certain situations, we give it extensive attention throughout this book.

In the future, we will be referring to distributions of scores. Whenever the term distribution appears, you should conjure up an image of a frequency distribution graph. The graph provides a picture showing exactly where the individual scores are located. To make this concept more concrete, you might find it useful to think of the graph as

**BOX 2.1 THE USE AND MISUSE OF GRAPHS**

Although graphs are intended to provide an accurate picture of a set of data, they can be used to exaggerate or misrepresent a set of scores. These misrepresentations generally result from failing to follow the basic rules for graph construction. The following example demonstrates how the same set of data can be presented in two entirely different ways by manipulating the structure of a graph.

For the past several years, the city has kept records of the number of homicides. The data are summarized as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Homicides</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>42</td>
</tr>
<tr>
<td>2006</td>
<td>44</td>
</tr>
<tr>
<td>2007</td>
<td>47</td>
</tr>
<tr>
<td>2008</td>
<td>49</td>
</tr>
</tbody>
</table>

These data are shown in two different graphs in Figure 2.9. In the first graph, we have exaggerated the height and started numbering the Y-axis at 40 rather than at zero. As a result, the graph seems to indicate a rapid rise in the number of homicides over the 4-year period. In the second graph, we have stretched out the X-axis and used zero as the starting point for the Y-axis. The result is a graph that shows little change in the homicide rate over the 4-year period.

Which graph is correct? The answer is that neither one is very good. Remember that the purpose of a graph is to provide an accurate display of the data. The first graph in Figure 2.9 exaggerates the differences between years, and the second graph conceals the differences. Some compromise is needed. Also note that in some cases a graph may not be the best way to display information. For these data, for example, showing the numbers in a table would be better than either graph.

**FIGURE 2.9**

Two graphs showing the number of homicides in a city over a 4-year period. Both graphs show exactly the same data. However, the first graph gives the appearance that the homicide rate is high and rising rapidly. The second graph gives the impression that the homicide rate is low and has not changed over the 4-year period.
showing a pile of individuals just like we showed a pile of blocks in Figure 2.3. For the population of IQ scores shown in Figure 2.8, the pile is highest at an IQ score around 100 because most people have average IQs. There are only a few individuals piled up at an IQ of 130; it must be lonely at the top.

2.4 THE SHAPE OF A FREQUENCY DISTRIBUTION

Rather than drawing a complete frequency distribution graph, researchers often simply describe a distribution by listing its characteristics. There are three characteristics that completely describe any distribution: shape, central tendency, and variability. In simple terms, central tendency measures where the center of the distribution is located. Variability tells whether the scores are spread over a wide range or are clustered together. Central tendency and variability will be covered in detail in Chapters 3 and 4. Technically, the shape of a distribution is defined by an equation that prescribes the exact relationship between each X and Y value on the graph. However, we will rely on a few less-precise terms that serve to describe the shape of most distributions.

Nearly all distributions can be classified as being either symmetrical or skewed.

DEFINITIONS

In a symmetrical distribution, it is possible to draw a vertical line through the middle so that one side of the distribution is a mirror image of the other (Figure 2.10).

In a skewed distribution, the scores tend to pile up toward one end of the scale and taper off gradually at the other end (see Figure 2.10).

The section where the scores taper off toward one end of a distribution is called the tail of the distribution.

FIGURE 2.10
Examples of different shapes for distributions.

Symmetrical distributions

Skewed distributions

Positive skew

Negative skew
A skewed distribution with the tail on the right-hand side is said to be positively skewed because the tail points toward the positive (above-zero) end of the X-axis. If the tail points to the left, the distribution is said to be negatively skewed (see Figure 2.10).

For a very difficult exam, most scores tend to be low, with only a few individuals earning high scores. This produces a positively skewed distribution. Similarly, a very easy exam tends to produce a negatively skewed distribution, with most of the students earning high scores and only a few with low values.

**LEARNING CHECK**

1. Sketch a frequency distribution histogram and a frequency distribution polygon for the data in the following table:

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

2. Describe the shape of the distribution in Exercise 1.

3. A researcher surveys a group of 400 college students and asks each person to identify his or her favorite movie from the preceding year. What type of graph should be used to show the distribution of responses?

4. A college reports that 20% of the registered students are older than 25. What is the shape of the distribution of ages for registered students?

**ANSWERS**

1. The graphs are shown in Figure 2.11.
2. The distribution is negatively skewed.
3. A bar graph is used for nominal data.
4. It would be positively skewed with most of the distribution around 18–21 and a few scores scattered at 25 and higher.

**FIGURE 2.11**

Answers to Learning Check Exercise 1.
1. The goal of descriptive statistics is to simplify the organization and presentation of data. One descriptive technique is to place the data in a frequency distribution table or graph that shows exactly how many individuals (or scores) are located in each category on the scale of measurement.

2. A frequency distribution table lists the categories that make up the scale of measurement (the X values) in one column. Beside each X value, in a second column, is the frequency or number of individuals in that category. The table may include a proportion column showing the relative frequency for each category:

\[ \text{proportion} = p = \frac{f}{n} \]

The table may include a percentage column showing the percentage associated with each X value:

\[ \text{percentage} = p(100) = \frac{f}{n} (100) \]

3. It is recommended that a frequency distribution table have a maximum of 10 to 15 rows to keep it simple. If the scores cover a range that is wider than this suggested maximum, it is customary to divide the range into sections called class intervals. These intervals are then listed in the frequency distribution table along with the frequency or number of individuals with scores in each interval. The result is called a grouped frequency distribution. The guidelines for constructing a grouped frequency distribution table are as follows:
   a. There should be about 10 intervals.
   b. The width of each interval should be a simple number (e.g., 2, 5, or 10).
   c. The bottom score in each interval should be a multiple of the width.
   d. All intervals should be the same width, and they should cover the range of scores with no gaps.

4. A frequency distribution graph lists scores on the horizontal axis and frequencies on the vertical axis. The type of graph used to display a distribution depends on the scale of measurement used. For interval or ratio scales, you should use a histogram or a polygon. For a histogram, a bar is drawn above each score so that the height of the bar corresponds to the frequency. Each bar extends to the real limits of the score, so that adjacent bars touch. For a polygon, a dot is placed above the midpoint of each score or class interval so that the height of the dot corresponds to the frequency; then lines are drawn to connect the dots. Bar graphs are used with nominal or ordinal scales. Bar graphs are similar to histograms except that gaps are left between adjacent bars.

5. Shape is one of the basic characteristics used to describe a distribution of scores. Most distributions can be classified as either symmetrical or skewed. A skewed distribution that tails off to the right is said to be positively skewed. If it tails off to the left, it is negatively skewed.

---

**KEY TERMS**

- frequency distribution (36)
- range (39)
- grouped frequency distribution (40)
- class interval (40)
- apparent limits (42)
- histogram (43)
- polygon (45)
- bar graph (46)
- relative frequency (47)
- symmetrical distribution (49)
- tail(s) of a distribution (49)
- positively skewed distribution (50)
- negatively skewed distribution (50)
RESOURCES

Book Companion Website: www.cengage.com/psychology/gravetter
You'll find a tutorial quiz and other learning exercises for Chapter 2.

WebAssign

Guided interactive tutorials, end-of-chapter problems, and related testbank items may be assigned online at WebAssign.

WebTutor

For those using WebTutor along with this book, there is a WebTutor section corresponding to this chapter. The WebTutor contains a brief summary of Chapter 2, hints for learning the new material and for avoiding common errors, and sample exam items including solutions.

SPSS

General instructions for using SPSS are presented in Appendix D. Following are detailed instructions for using SPSS to produce Frequency Distribution Tables or Graphs.

Frequency Distribution Tables

Data Entry

1. Enter all the scores in one column of the data editor, probably VAR00001.

Data Analysis

1. Click Analyze on the tool bar, select Descriptive Statistics, and click on Frequencies.
2. Highlight the column label for the set of scores (VAR00001) in the left box and click the arrow to move it into the Variable box.
3. Be sure that the option to Display Frequency Table is selected.
4. Click OK.

SPSS Output

The frequency distribution table will list the score values in a column from smallest to largest, with the percentage and cumulative percentage also listed for each score. Score values that do not occur (zero frequencies) are not included in the table, and the program does not group scores into class intervals (all values are listed).

Frequency Distribution Histograms or Bar Graphs

Data Entry

1. Enter all the scores in one column of the data editor, probably VAR00001.
Data Analysis

1. Click Analyze on the tool bar, select Descriptive Statistics, and click on Frequencies.
2. Highlight the column label for the set of scores (VAR00001) in the left box and click the arrow to move it into the Variable box.
3. Click Charts.
4. Select either Bar Graphs or Histogram.
5. Click Continue.
6. Click OK.

SPSS Output

After a brief delay, SPSS will display a frequency distribution table and a graph. Note that SPSS often produces a histogram that groups the scores in unpredictable intervals. A bar graph usually produces a clearer picture of the actual frequency associated with each score.

FOCUS ON PROBLEM SOLVING

1. The reason for constructing frequency distributions is to put a disorganized set of raw data into a comprehensible, organized format. Because several different types of frequency distribution tables and graphs are available, one problem is deciding which type should be used. Tables have the advantage of being easier to construct, but graphs generally give a better picture of the data and are easier to understand.
   To help you decide which type of frequency distribution is best, consider the following points:
   a. What is the range of scores? With a wide range, you need to group the scores into class intervals.
   b. What is the scale of measurement? With an interval or a ratio scale, you can use a polygon or a histogram. With a nominal or an ordinal scale, you must use a bar graph.

2. When using a grouped frequency distribution table, a common mistake is to calculate the interval width by using the highest and lowest values that define each interval. For example, some students are tricked into thinking that an interval identified as 20–24 is only 4 points wide. To determine the correct interval width, you can
   a. Count the individual scores in the interval. For this example, the scores are 20, 21, 22, 23, and 24 for a total of 5 values. Thus, the interval width is 5 points.
   b. Use the real limits to determine the real width of the interval. For example, an interval identified as 20–24 has a lower real limit of 19.5 and an upper real limit of 24.5 (halfway to the next score). Using the real limits, the interval width is

24.5 − 19.5 = 5 points

DEMONSTRATION 2.1

A GROUPED FREQUENCY DISTRIBUTION TABLE

For the following set of \( N = 20 \) scores, construct a grouped frequency distribution table using an interval width of 5 points.
The scores are:

14, 8, 27, 16, 10, 22, 9, 13, 16, 12,
10, 9, 15, 17, 6, 14, 11, 18, 14, 11
CHAPTER 2  FREQUENCY DISTRIBUTIONS

**STEP 1** Set up the class intervals.

The largest score in this distribution is \( X = 27 \), and the lowest is \( X = 6 \). Therefore, a frequency distribution table for these data would have 22 rows and would be too large. A grouped frequency distribution table would be better. We have asked specifically for an interval width of 5 points, and the resulting table has five rows.

<table>
<thead>
<tr>
<th>( X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–29</td>
</tr>
<tr>
<td>20–24</td>
</tr>
<tr>
<td>15–19</td>
</tr>
<tr>
<td>10–14</td>
</tr>
<tr>
<td>5–9</td>
</tr>
</tbody>
</table>

Remember that the interval width is determined by the real limits of the interval. For example, the class interval 25–29 has an upper real limit of 29.5 and a lower real limit of 24.5. The difference between these two values is the width of the interval—namely, 5.

**STEP 2** Determine the frequencies for each interval.

Examine the scores, and count how many fall into the class interval of 25–29. Cross out each score that you have already counted. Record the frequency for this class interval. Now repeat this process for the remaining intervals. The result is the following table:

<table>
<thead>
<tr>
<th>( X )</th>
<th>( f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–29</td>
<td>1</td>
</tr>
<tr>
<td>20–24</td>
<td>1</td>
</tr>
<tr>
<td>15–19</td>
<td>5</td>
</tr>
<tr>
<td>10–14</td>
<td>9</td>
</tr>
<tr>
<td>5–9</td>
<td>4</td>
</tr>
</tbody>
</table>

**PROBLEMS**

1. Place the following sample of \( n = 20 \) scores in a frequency distribution table.

\[
\begin{align*}
6, & 9, 9, 10, 8, 9, 4, 7, 10, 9 \\
5, & 8, 10, 6, 9, 6, 8, 8, 7, 9
\end{align*}
\]

2. Construct a frequency distribution table for the following set of scores. Include columns for proportion and percentage in your table.

Scores: 5, 7, 8, 4, 7, 9, 6, 6, 5, 3
9, 6, 4, 7, 7, 8, 6, 7, 8, 5

3. Find each value requested for the distribution of scores in the following table.

   a. \( n \)
   b. \( \Sigma X \)
   c. \( \Sigma X^2 \)

<table>
<thead>
<tr>
<th>( X )</th>
<th>( f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

4. Find each value requested for the distribution of scores in the following table.

   a. \( n \)
   b. \( \Sigma X \)
c. $\sum x^2$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

5. For the following scores, the smallest value is $x = 8$ and the largest value is $x = 29$. Place the scores in a grouped frequency distribution table.
   a. using an interval width of 2 points.
   b. using an interval width of 5 points.
   24, 19, 23, 10, 25, 27, 22, 26
   25, 20, 8, 24, 29, 21, 24, 13
   23, 27, 24, 16, 22, 18, 26, 25

6. The following scores are the ages for a random sample of $n = 30$ drivers who were issued speeding tickets in New York during 2008. Determine the best interval width and place the scores in a grouped frequency distribution table. From looking at your table, does it appear that tickets are issued equally across age groups?
   17, 30, 45, 20, 39, 53, 28, 19, 24, 21, 34, 38, 22, 29, 64, 22, 44, 36, 16, 56, 20, 23, 58, 32, 25, 28, 22, 51, 26, 43

7. For each of the following samples, determine the interval width that is most appropriate for a grouped frequency distribution and identify the approximate number of intervals needed to cover the range of scores.
   a. Sample scores range from $X = 24$ to $X = 41$
   b. Sample scores range from $X = 46$ to $X = 103$
   c. Sample scores range from $X = 46$ to $X = 133$

8. Under what circumstances should you use a grouped frequency distribution instead of a regular frequency distribution?

9. What information can you obtain about the scores in a regular frequency distribution table that is not available from a grouped table?

10. Describe the difference in appearance between a bar graph and a histogram and describe the circumstances in which each type of graph is used.

11. For the following set of quiz scores:
    3, 5, 4, 6, 2, 3, 4, 1, 4, 3
    7, 7, 3, 4, 5, 8, 2, 4, 7, 19
   a. Construct a frequency distribution table to organize the scores.
   b. Draw a frequency distribution histogram for these data.

12. Sketch a histogram and a polygon showing the distribution of scores presented in the following table:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

13. Find each of the following values for the distribution of scores shown in the frequency distribution polygon.
   a. $N$
   b. $\Sigma x$

14. A survey given to a sample of 200 college students contained questions about the following variables. For each variable, identify the kind of graph that should be used to display the distribution of scores (histogram, polygon, or bar graph).
   a. number of pizzas consumed during the previous week
   b. size of T-shirt worn (S, M, L, XL)
   c. gender (male/female)
   d. grade point average for the previous semester
   e. college class (freshman, sophomore, junior, senior)

15. Each year the college gives away T-shirts to new students during freshman orientation. The students are allowed to pick the shirt sizes that they want. To determine how many of each size shirt they should order, college officials look at the distribution from
last year. The following table shows the distribution of shirt sizes selected last year.

<table>
<thead>
<tr>
<th>Size</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>27</td>
</tr>
<tr>
<td>M</td>
<td>48</td>
</tr>
<tr>
<td>L</td>
<td>136</td>
</tr>
<tr>
<td>XL</td>
<td>120</td>
</tr>
<tr>
<td>XXL</td>
<td>39</td>
</tr>
</tbody>
</table>

20. For the following set of scores:

\[ 3, 7, 6, 5, 5, 9, 6, 4, 6, 8 \]
\[ 10, 2, 7, 4, 9, 5, 6, 3, 8 \]

a. Construct a frequency distribution table.
b. Sketch a polygon showing the distribution.
c. Describe the distribution using the following characteristics:
   (1) What is the shape of the distribution?
   (2) What score best identifies the center (average) for the distribution?
   (3) Are the scores clustered together, or are they spread out across the scale?

21. Fowler and Christakis (2008) report that personal happiness tends to be associated with having a social network including many other happy friends. To test this claim, a researcher obtains a sample of \( n = 16 \) adults who claim to be happy people and a similar sample of \( n = 16 \) adults who describe themselves as neutral or unhappy. Each individual is then asked to identify the number of their close friends whom they consider to be happy people. The scores are as follows:

Happy: \[ 8, 7, 4, 10, 6, 6, 8, 9, 8, 8, 7, 5, 6, 9, 9, 8, 9 \]

Unhappy: \[ 5, 8, 4, 6, 6, 7, 9, 6, 2, 8, 5, 6, 4, 7, 5, 6 \]

Sketch a polygon showing the frequency distribution for the happy people. In the same graph, sketch a polygon for the unhappy people. (Use two different colors, or use a solid line for one polygon and a dashed line for the other.) Does one group seem to have more happy friends?

22. Schmidt (1994) conducted a series of experiments examining the effects of humor on memory. In one study, participants were shown a list of sentences, of which half were humorous and half were nonhumorous. A humorous example is, “If at first you don’t succeed, you are probably not related to the boss.” Other participants would see a nonhumorous version of this sentence, such as “People who are related to the boss often succeed the very first time.”

Schmidt then measured the number of each type of sentence recalled by each participant. The following scores are similar to the results obtained in the study.
<table>
<thead>
<tr>
<th>Number of Sentences Recalled</th>
<th>Humorous sentences</th>
<th>Nonhumorous sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5 2 4</td>
<td>5 2 4 2</td>
</tr>
<tr>
<td>6</td>
<td>7 6 6</td>
<td>2 3 1 6</td>
</tr>
<tr>
<td>2</td>
<td>5 4 3</td>
<td>3 2 3 3</td>
</tr>
<tr>
<td>1</td>
<td>3 5 5</td>
<td>4 1 5 3</td>
</tr>
</tbody>
</table>

a. Identify the independent variable and the dependent variable for this experiment.

b. Sketch a polygon showing the frequency distribution for the humorous sentences. In the same graph, sketch a polygon for the nonhumorous sentences. (Use two different colors, or use a solid line for one polygon and a dashed line for the other.) Does it appear that humor has an influence on memory?