Frequency Distributions

2

LEARNING OBJECTIVES

- Make a frequency distribution for a set of data.
- Decide if a number is discrete or continuous.
- Choose and make the appropriate graph for a frequency distribution.
- Describe modality, skewness, and kurtosis for a frequency distribution.
- Make a stem and leaf display.

CHAPTER OVERVIEW

The purpose of statistics is data reduction, taking a mass of numbers and reducing them in some way to bring order to them. Chapter 1 showed how organizing binge drinking rates for 18- to 25-year-olds in the United States—alphabetically or from low to high—made them easier to understand. That just involved reorganizing the data, not reducing it. Chapter 2 covers some data reduction techniques, making tables and graphs. It also introduces another way to think about numbers, whether the number is continuous or discrete, and talks about the "shapes" data can take.

2.1 Frequency Distributions

Frequency distributions can be made for nominal-, ordinal-, interval-, or ratio-level data. A frequency distribution is an intuitive way to organize and reduce data. A **frequency distribution** is simply a count of how often the values of a variable occur in a set of data. For example, to tell someone how many boys and girls are in a class is to make a frequency distribution.

Ungrouped vs. Grouped Frequency Distributions

There are two different types of frequency distribution tables, ungrouped and grouped. An **ungrouped frequency distribution** table is a count of how often each value of a variable occurs in a data set. In a **grouped frequency distribution** table, the frequency counts are for adjacent groupings of values, or intervals, of the variable.

Ungrouped frequency distributions are used when the values a variable can take are limited. For example, if students were surveyed about how many children were in their families, a limited number of responses would exist. Students would most commonly report that there were one, two, or three kids in their families. Almost no one would report ten or more kids. An ungrouped frequency distribution for data like these would be compact, taking up maybe a half-dozen lines on one page, and could be viewed easily.

Now if the same students were surveyed about the size of their high school graduating classes, it would be a very different frequency distribution. The survey would yield answers ranging from 1 (homeschooled) to 500 or more students. If one listed each value separately, it would take many lines and could run to multiple pages. This would be hard to view and not a good summary of the data. In this case, it would make sense to group answers together in intervals (fewer than 100 in the class, 100 to 199 in the class, etc.) to make a more compact presentation.

Grouped frequency distributions should be used when the variable has a large number of values and it is acceptable to lose information by collapsing the values into intervals. If the variable has a large number of values but it is important to retain information about all the unique values, then one should use an ungrouped frequency distribution.

Table 2.1 shows both an ungrouped frequency distribution and a grouped frequency distribution for the binge drinking data from Chapter 1. Note the size of the ungrouped frequency distribution and the many gaps in it. There is too much detail to get a clear picture of the data. Compare that to the compactness of the grouped frequency distribution, which reduces the data to a greater degree and does a better job summarizing and describing it.

Ungrouped Frequency Distributions

Here are some data regarding how many children are in the families of 31 students. Nine of the 31 reported being in one-child families, 14 families had two children, 5 had three, 2 had four, and 1 had six. An ungrouped frequency distribution table for these data can be seen in Table 2.2 on page 42.

There are a number of things to note about Table 2.2:

- This is a basic, no-frills frequency distribution. It just provides the values that the variable takes and how often each value occurs. That's it.
- There is a title! All tables need to have titles that clearly describe the information the table contains.
- The columns are labeled.
- The abbreviation for frequency is *f*.
- The table is "upside down," with the largest value of the variable (six children per family) at the top and the smallest value (one child per family) at the bottom. Why the table is arranged this way will become clear in a moment when cumulative frequencies are introduced.
- There is no value in the table that indicates how many students were in the class, so that information was placed in the title. The sample size, *N*, is commonly reported in tables and graphs.
- There was one value (five children per family) that did not exist in the class. But, this is included anyway with a frequency of zero. Include zero frequency values because that shows breaks in the data set.

TABLE 2.1 Comparison of Ungrouped and Grouped Frequency Distributions for Data Showing Percentage of 18- to 25-Year-Olds per State Who Engaged in Binge Drinking During the Past 30 Days

Ungrouped Frequency Distribution

Percentage of 18- to 25-Year-Olds Per State Who Engaged in Binge Drinking During Past Month	Number of States That Had That Percentage of Binge Drinking 18- to 25-Year-Olds
54	1
53	
52	
51	
50	
49	2
48	1
47	1
46	2
45	4
44	4
43	1
42	2
41	7
40	1
39	6
38	3
37	2
36	4
35	2
34	
33	2
32	2
31	1
30	1
29	
28	
27	
26	1

Grouped Frequency Distribution

Range for Percentage of 18- to 25-Year-Olds Per State Who Engaged in Binge Drinking During Past Month	Number of States That Had a Percentage of Binge Drinking 18- to 25-Year-Olds in That Range
50–54	1
45–49	10
40–44	15
35–39	17
30–34	6
25–29	1

Both of these frequency distributions reduce the data shown in Table 1.3. Which one gives a clearer picture of the binge drinking rates, by state, of 18- to 25-year-olds? The ungrouped version shows the whole range of scores but has a lot of empty categories. The much more compact grouped version shows the big picture but sacrifices detail.

TABLE 2.2 A Basic Ungrouped Frequency Distribution for Number of Children in Families of 31 Students in a Class

Number of Children in Family	Frequency (*f*)
6	1
5	0
4	2
3	5
2	14
1	9

Note: This is the simplest version of a frequency distribution. The only information it contains is the frequency with which each value occurs.

Table 2.2 is a basic ungrouped frequency distribution. Table 2.3 is one with more bells and whistles. This more complex ungrouped frequency distribution has three new columns. The first, **cumulative frequency,** tells how many cases in a data set have a given value *or* a lower value. For example, there are 23 people who have two *or fewer* children in their family and 28 who have three *or fewer.*

TABLE 2.3 Ungrouped Frequency Distribution for Number of Children in Families of 31 Students in a Class

Number of Children in Family	Frequency (*f*)	Cumulative Frequency (f_c)	Percentage (%)	Cumulative Percentage ($\%_c$)
6	1	31	3.23	100.00
5	0	30	0.00	96.77
4	2	30	6.45	96.77
3	5	28	16.13	90.32
2	14	23	45.16	74.19
1	9	9	29.03	29.03

Note: This ungrouped frequency distribution contains more information than the basic version in Table 2.2.

A cumulative frequency, abbreviated f_c, is calculated by adding up all the frequencies at or below a given row. It is easier to visualize this than to make it into a formula. Look at Figure 2.1 and note how the cumulative frequencies stair-step up. For the first row, the frequency and the cumulative frequency are the same, 9. Moving up one step, add the frequency at this level, 14, to the cumulative frequency from the level below, 9, to get the cumulative frequency of 23 for this

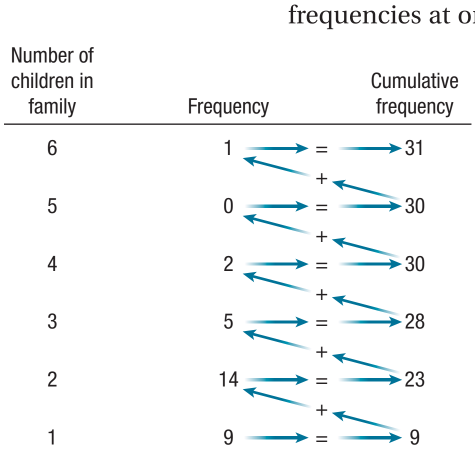


Figure 2.1 Calculating Cumulative Frequencies for a Frequency Distribution One can calculate cumulative frequencies by "stair-stepping" up. The cumulative frequency in the bottom row is the same as the frequency for that row. Then add the frequency from the row above to find its cumulative frequency. Continue the process until one reaches the final row. Note that the cumulative frequency in the top row is the same as the number of cases in the data set.

level. Then repeat the process—add the frequency at the new level to the cumulative frequency from one step down—to get the next cumulative frequency.

There are a few other things one should know about cumulative frequencies:

- Frequency distributions are organized upside down, with the biggest value the variable can take on the top row. This is because of cumulative frequency, the number of cases at or *below* a given value. In this table, the row for six children in the family is the top row and the row for one child in the family is the bottom row.
- The cumulative frequency for the top row should be the same as the total number of cases, N, in the data set. If they don't match, something is wrong.
- Cumulative frequencies can only be calculated for data that have an order, data where the numbers tell direction. This means that cumulative frequencies can be calculated for ordinal-, interval-, or ratio-level data, but not for nominal data.
- If one has nominal data, one needs to organize the data in some logical fashion to fit one's intended purpose. For example, one could make a frequency distribution for the nominal variable of choice of college major. To draw attention to the relative popularity of majors, it could be organized by ascending or descending frequency.

Cumulative frequencies can only be calculated for data that have an order, data where the numbers tell direction.

The other new columns in Table 2.3 take information that is already there, frequency and cumulative frequency, and present it as percentages. Percentages are a way of transforming scores to put them in context. To see how this works, imagine that a psychologist said that two of the clients he treated in the last year experienced major depression. Does he specialize in the treatment of depression? The answer depends on how many total clients he treated in the last year. If he offered therapy to four patients, then half of them, or 50%, were depressed. If he treated 100 cases, then only 2% were depressed. Whether the psychologist is a depression specialist depends on what percentage of his practice is devoted to treating that illness. Percentages put scores in context.

The percentage column, abbreviated %, takes the information in the frequency column and turns it into percentages. Equation 2.1 shows that this is done by dividing a frequency by the total number of cases in the data set and then multiplying the quotient by 100.

Equation 2.1 Formula for Calculating Frequency Percentage (%) for a Frequency Distribution

$$\% = \frac{f}{N} \times 100$$

where % = frequency percentage
f = frequency
N = total number of cases

For example, in the third row where the frequency is 5, one would calculate

$$\begin{aligned}\% &= \frac{5}{31} \times 100 \\ &= 0.1613 \times 100 \\ &= 16.13\%\end{aligned}$$

This means, in plain language, that 16.13% of the students come from families with three children.

The final column, abbreviated $\%_c$, gives the **cumulative percentage,** the percentage of cases with scores at or below a given level. The cumulative percentage is a restatement of the information in the cumulative frequency column in percentage format. For example, in the second row from the bottom of Table 2.3, it is readily apparent that just over 74% of the students come from one-child or two-child families.

Equation 2.2 Formula for Calculating Cumulative Percentage ($\%_c$) for a Frequency Distribution

$$\%_c = \frac{f_c}{N} \times 100$$

where $\%_c$ = cumulative percentage
f_c = cumulative frequency
N = total number of cases

The formula for cumulative percentage is given in Equation 2.2. To find $\%_c$ for a given row in a frequency distribution table, one divides the row's cumulative frequency by the total number of cases and then multiplies the quotient by 100. In the second to last row of Table 2.3, where the cumulative frequency is 23, one would calculate

$$\begin{aligned}\%_c &= \frac{23}{31} \times 100 \\ &= 0.7419 \times 100 \\ &= 74.19\end{aligned}$$

Note that the cumulative percentage for the top row should equal 100%.

Grouped Frequency Distributions

Ungrouped frequency distributions work well in two conditions: (1) when the variable takes a limited set of values, or (2) when one wants to document all the values a variable can take. If one were classifying semester status of students in college, there is a limited number of options (1st, 2nd, 3rd, etc.) and an ungrouped frequency distribution would work well. There are many more options for the number of credit hours a student has completed, and an ungrouped frequency distribution for this variable would only make sense if it were important to keep track of how many students had 15 hours vs. 16 hours vs. 17 hours, and so forth.

When dealing with a variable that has a large range, like number of credit hours, a grouped frequency distribution makes more sense. In a grouped frequency distribution, one finds the frequency with which a variable occurs over a range of values. In the credit hours example, one might count how many students have completed from 0 to 14 hours, or from 15 to 29 hours, and so forth. These ranges are called intervals or bins and are abbreviated with a lowercase *i*.

Grouped frequency distributions work best when there is an order to the values a variable can take, that is, when the variable is measured at the ordinal, interval, or ratio level. Nominal data can be grouped if there is some logical categorization. For example, imagine that a psychologist collected detailed information about the diagnoses of her patients—whether they had major depression, dysthymia, bipolar disorder, obsessive-compulsive disorder, phobias, generalized anxiety disorder, alcoholism, heroin addiction. These responses could be grouped into categories of mood disorders, anxiety disorders, and substance abuse disorders.

TABLE 2.4 Property Crime Rates for the 50 States

State	Rate	State	Rate	State	Rate
South Dakota	17	Wisconsin	28	Kansas	37
North Dakota	19	Wyoming	29	New Mexico	37
New Hampshire	19	Illinois	29	Missouri	37
New York	20	Colorado	30	Nevada	38
New Jersey	22	California	30	Georgia	39
Idaho	22	Minnesota	30	Arkansas	40
Vermont	23	Michigan	31	Alabama	40
Pennsylvania	24	Nebraska	32	Washington	40
Massachusetts	24	Mississippi	32	Louisiana	41
Connecticut	24	Delaware	34	North Carolina	41
Maine	24	Alaska	34	Tennessee	41
Virginia	25	Indiana	34	Florida	41
Kentucky	25	Maryland	34	Texas	41
West Virginia	25	Ohio	35	Hawaii	42
Iowa	26	Utah	35	South Carolina	43
Rhode Island	26	Oregon	35	Arizona	44
Montana	28	Oklahoma	35		

Note: Property crime rates are reported as the number of crimes that occur per every 1,000 people who live in the state. Here, the data are organized in order from low to high.

Source: Statistical Abstract of the United States.

For variables that are measured at the ordinal level or higher, the first step is to decide how many intervals to include in a grouped frequency distribution. There needs to be a balance between the amount of detail presented and the number of intervals. There shouldn't be so few intervals that one can't see important details in the data set, and there shouldn't be so many intervals that the big picture is lost in the details. A rule of thumb is to use psychology's magic number, 7 ± 2, and have from five to nine intervals. But, that's just a rule of thumb—if fewer than five intervals or more than nine intervals do a better job of communication, that's fine. The other thing to keep in mind is convention. It is common to make intervals that are 5, 10, 20, 25, or 100 units wide.

Table 2.4 displays data we'll use to make a grouped frequency distribution. These numbers represent the number of property crimes that occur in a state for every thousand people who live in that state. They are organized from low (17 crimes per 1,000) to high (44 crimes per 1,000) to make it easy to construct a grouped frequency distribution.

The first task is to decide how many intervals to include. One needs to have intervals that will capture all data and that don't overlap. One option is to have 10-point-wide intervals—one for the values in the 10s, one for the 20s, the 30s, and the 40s. But that means only four intervals and so few intervals would lose too much detail. Instead, it makes sense here to opt for 5-point-wide intervals. Note that the first interval starts at 15, not 17, because starting an interval at a multiple of 5 is a convention commonly followed. The first interval is 15–19. (If you don't believe it is five points wide, count it off on your fingers.)

There's one other thing to mention—all intervals should be the same width so that the frequency in one interval can be compared to the frequency in another. So, there will be six intervals, ranging from 15–19 on the bottom row to 40–44 on the top.

TABLE 2.5 Grouped Frequency Distribution for State Property Crime Rates per 1,000 Population (Interval Width = 5)

Crime Rate Interval	Midpoint (m)	Frequency (f)	Cumulative Frequency (f_c)	Percentage (%)	Cumulative Percentage ($\%_c$)
40–44	42.00	11	50	22.00	100.00
35–39	37.00	9	39	18.00	78.00
30–34	32.00	10	30	20.00	60.00
25–29	27.00	9	20	18.00	40.00
20–24	22.00	8	11	16.00	22.00
15–19	17.00	3	3	6.00	6.00

Note: This table organizes the data from all 50 states in Table 2.4 into six 5-point intervals, giving an overview of the data.

Table 2.5 shows what a completed grouped frequency distribution looks like for these state crime data.

As with an ungrouped frequency distribution, a grouped frequency distribution has a title, labeled columns, and is upside down. A grouped frequency distribution also presents the same information—frequency, cumulative frequency, percentage, and cumulative percentage. This grouped frequency distribution does have one new column, m, which gives information about the *midpoint* of each interval. The **midpoint** is just what it sounds like, the middle point of the interval. One can verify that the midpoint of the first interval is 17.00 by counting through the interval, 15–19, on one's fingers. Or, find the average of 15 and 19, the endpoints of the interval

$$\frac{15 + 19}{2} = 17.00$$

There are two reasons for having a midpoint. First, the midpoint is, in a sense, a summary of the interval. Imagine a compass drawn as a circle with only four headings—north, east, south, and west (see Figure 2.2). North is at the top, labeled 360°, east is to the right, labeled 90°, and so on. If Sharon were heading off to the right at exactly 90°, one would say that she was heading east. But when else would one say she was heading east? If the only options were to say a person was heading north, east, south, or west, one would say "east" when that person was heading anywhere from 45° to 135°. East is the midpoint of that interval $\left(\frac{45 + 135}{2} = 90.00\right)$, a summary of the direction for all the points in that interval.

There's another reason for having a midpoint. Suppose a grouped frequency distribution, like that in Table 2.5, was the only information one had. There was no access to the original data. One would know that there are three cases in the 15–19 interval, but their specific values wouldn't be known. Were they all 15s? All 19s? Spread throughout the interval? Statisticians have developed the convention that when the exact value for a case is unknown, it is assigned the value of the midpoint of the interval within which it falls.

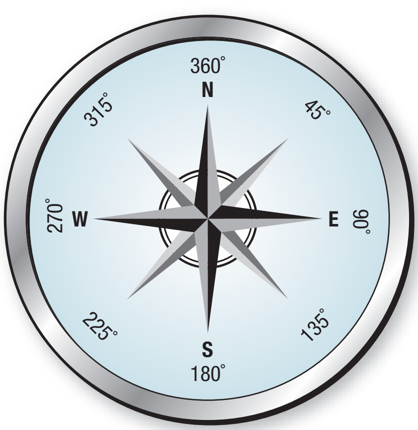


Figure 2.2 Four Cardinal Compass Headings This figure shows the degrees associated with the four main compass headings—north, east, south, and west—as well as the degrees halfway between them. Though a heading of east is exactly 90°, we would say a person is heading east anywhere from 45° to 135°. "East" is the midpoint of that interval.

A lot of material was covered in this section, so here are some summaries. Figure 2.3 is a flowchart that leads one through the process of deciding whether to make an ungrouped or a grouped frequency distribution. Table 2.6 summarizes what material to include when making a table.

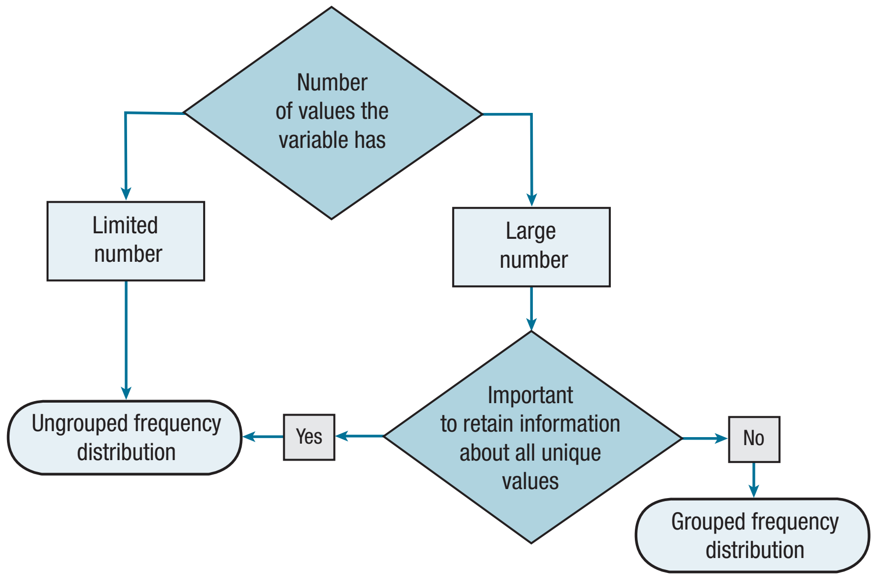


Figure 2.3 How to Choose: Deciding Whether to Make an Ungrouped Frequency Distribution or a Grouped Frequency Distribution for an Ordinal-, Interval-, or Ratio-Level Variable This flowchart leads one through the decision process of whether to make an ungrouped frequency distribution or a grouped frequency distribution for a variable that is measured at a level higher than nominal.

TABLE 2.6 How to Choose: Making a Frequency Distribution Table

Step 1	Decide whether to make a grouped or ungrouped frequency distribution: • Grouped, if the variable takes a wide range of values. • Ungrouped, if important to maintain information about all values.
Step 2	If grouped, decide the width and number of intervals: • Aim to have 5–9 intervals. • Make sure the intervals capture all values, have the same width, and don't overlap.
Step 3	Organize the data: • For ordinal, interval, or ratio variables, organize the table "upside down," with the largest value/interval on the top row and the rest of the values following in descending order. • For nominal variables, apply some order to the values (e.g., alphabetical, like categories together, ascending or descending frequencies).
Step 4	Decide what information to include in the table: • At a minimum, the values of the variable and their respective frequencies should be included. • Optional: Cumulative frequencies (not for nominal variables), percentage, cumulative percentage, interval midpoints.
Step 5	Communicate clearly: • Have a descriptive title. • Label all columns. • If abbreviations are used, include a note of explanation.

Worked Example 2.1

For practice making a frequency distribution table, imagine a school district that wanted to get a better picture of the intelligence of its students. It hired a psychologist to administer IQ tests to a random sample of sixth-grade students. Now the district needs a summary of the results.

Table 2.7 shows the raw data for the students tested. The first thing to note is that there is a lot of data—four columns of 15 and one of 8, so $N = 68$. The second thing to note is that a wide range of IQ scores exists, from the 60s to the 150s. Given the wide range of scores and given that it doesn't seem necessary to maintain information about all the unique scores, a grouped frequency distribution is a more sensible option than an ungrouped frequency distribution to summarize the data.

TABLE 2.7 Sample Data: IQ Test Results for 68 Sixth Graders

74	109	100	76	101	134	97	67
126	105	134	95	90	152	142	101
111	111	147	119	108	89	122	90
121	105	109	119	129	128	109	102
73	98	106	82	75	68	98	115
95	102	97	85	84	80	122	
111	72	85	148	111	112	136	
92	94	105	93	116	88	90	
80	107	118	79	103	94	128	

Note: These IQ test results are in no particular order.

The first step in making a grouped frequency distribution is organizing the data. Table 2.8 shows the data arranged in ascending order.

TABLE 2.8 Sixth-Grade IQ Data Arranged in Order

67	80	90	98	105	111	121	136
68	82	92	98	105	111	122	142
72	84	93	100	106	111	122	147
73	85	94	101	107	112	126	148
74	85	94	101	108	115	128	152
75	88	95	102	109	116	128	
76	89	95	102	109	118	129	
79	90	97	103	109	119	134	
80	90	97	105	111	119	134	

Note: These are the same data as in Table 2.7, just arranged in ascending order of IQ score.

The next decision is how many intervals to have and/or how wide they should be. The range of scores is from 67 to 152. Grouping scores by tens into 60s, 70s, 80s, and so on seems like a reasonable approach. That would mean having 10 intervals, more than the five-to-nine-intervals rule of thumb suggests. But, it is not many more and having an interval width (10) that is familiar is important.

Next, one needs to decide what information to put into the frequency distribution. The bare minimum is the interval and the frequency, but tables are more useful if they also include cumulative frequency, percentage, and cumulative percentage. Table 2.9 shows a table, empty except for title, column labels, and interval information. Note that the intervals appear in the reverse order they did in Table 2.8. And, note that there are enough non-overlapping intervals so that each value falls in one, and only one, interval.

TABLE 2.9 Template for Grouped Frequency Distribution for IQ Scores of a Random Sample of Sixth Graders (Interval Width = 10)

IQ Score Interval	Midpoint (m)	Frequency (f)	Cumulative Frequency (f_c)	Percentage (%)	Cumulative Percentage ($\%_c$)
150–159					
140–149					
130–139					
120–129					
110–119					
100–109					
90–99					
80–89					
70–79					
60–69					

Note: This empty table is ready to be filled with data. Note that there is already a clear title, all columns are labeled, the interval with the lowest scores is on the bottom row, the intervals are all the same width, and the intervals are non-overlapping.

The completed table, Table 2.10, will provide the school district with a good descriptive summary of the intelligence of its sixth graders. It shows the range of scores, from the 60s to the 150s, and the most common scores, from 90 to 119.

TABLE 2.10 Grouped Frequency Distribution for IQ Scores of a Random Sample of Sixth Graders (Interval Width = 10)

IQ Score Interval	Midpoint (m)	Frequency (f)	Cumulative Frequency (f_c)	Percentage (%)	Cumulative Percentage ($\%_c$)
150–159	154.5	1	68	1.47	100.00
140–149	144.5	3	67	4.41	98.53
130–139	134.5	3	64	4.41	94.12
120–129	124.5	7	61	10.29	89.71
110–119	114.5	10	54	14.71	79.41
100–109	104.5	15	44	22.06	64.71
90–99	94.5	13	29	19.12	42.65
80–89	84.5	8	16	11.76	23.53
70–79	74.5	6	8	8.82	11.76
60–69	64.5	2	2	2.94	2.94

Note: This grouped frequency distribution makes it easier to see the range of scores and what the most common scores are.

Practice Problems 2.1

Review Your Knowledge

2.01 Under what conditions should an ungrouped frequency distribution be made? A grouped frequency distribution?

2.02 What information does a cumulative frequency provide?

2.03 How should a frequency distribution be organized for a nominal variable like type of religion?

2.04 How many intervals should a grouped frequency distribution have?

2.05 What is the formula to calculate a midpoint for an interval in a grouped frequency distribution?

Apply Your Knowledge

2.06 An exercise physiologist has first-year college students exercise and then measures how many minutes it takes for their heart rates to return to normal. Make an ungrouped frequency distribution that shows frequency, cumulative frequency, percentage, and cumulative percentage for the following data:

1, 3, 7, 2, 8, 12, 11, 3, 5, 6, 7, 4, 14, 8, 2, 3, 5, 8, 11, 10, 9, 8, 4, 3, 2, 3, 4, 2, 6, and 7

2.07 Below are final grades, in order, from an upper-level psychology class. Make a grouped frequency distribution to show the distribution of grades. Use an interval width of 10 and start the lowest interval at 50. Be sure to include midpoint, frequency, cumulative frequency, percentage, and cumulative percentage for the data:

53, 54, 56, 59, 60, 60, 60, 62, 63, 63, 66, 67, 67, 70, 70, 70, 71, 72, 73, 74, 74, 75, 75, 77, 77, 77, 78, 79, 79, 80, 80, 81, 81, 81, 81, 81, 83, 83, 84, 84, 85, 85, 85, 85, 86, 86, 87, 87, 88, 91, 91, 92, 93, 94, 96, 98, 98, 99

2.2 Discrete Numbers and Continuous Numbers

Chapter 1 covered levels of measurement, classifying numbers as nominal, ordinal, interval, or ratio. Now let's learn another way to classify numbers, as discrete or continuous. Knowing whether a number is discrete or continuous will be important in the next section of this chapter, which addresses how to graph a frequency distribution.

Discrete numbers answer the question "How many?" For example, the answers to questions like how many siblings one has, how many neurons are in a spinal cord, or how many jeans are in a closet are all discrete numbers. Discrete numbers take whole number values only and have no "in-between," or fractional, values. No matter how raggedy a person's favorite pair of jeans is, it's still one pair of jeans. One would never say that a pair of jeans with a lot of holes is 0.79 of a pair of jeans. Nominal- and ordinal-level numbers are always discrete. Sometimes, interval- and ratio-level numbers are discrete.

Because fractional values for continuous numbers represent distance, continuous numbers are always interval- or ratio-level numbers. But not all interval- or ratio-level numbers are continuous.

Continuous numbers answer the question "How much?" For example, the answers to questions like how much aggression a person has, how much intelligence one has, or how much a person weighs would be continuous numbers. Continuous numbers can take on values between whole numbers, that is, they may have fractional values. The number of decimal places reported for continuous numbers, how specific they are, depends on the measuring instrument used. More precise measuring instruments allow attributes to be measured more exactly.

Because fractional values for continuous numbers represent distance, continuous numbers are always interval- or ratio-level numbers. Keep in mind, though,

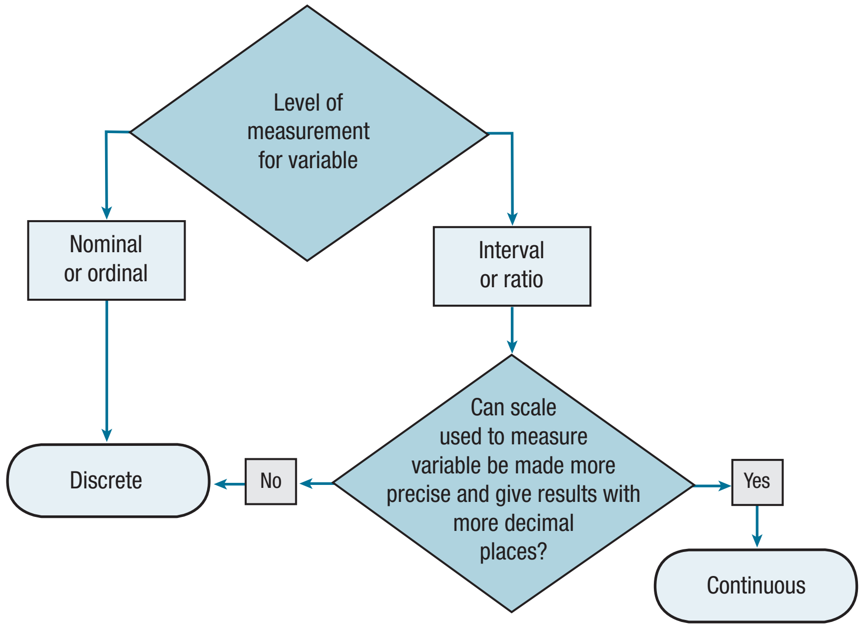


Figure 2.4 How to Choose: Continuous Numbers vs. Discrete Numbers Start by using Figure 1.10 on page 20 to determine the variable's level of measurement. Then, depending on whether the variable is (a) nominal or ordinal or (b) interval or ratio, follow the flowchart to determine if the variable is continuous or discrete.

that not all interval- or ratio-level numbers are continuous. Figure 2.4 is a flowchart that leads one through the process of determining whether a number is discrete or continuous.

A Common Question

Q If a number, like how many children in a family, is discrete and only takes whole number values, why is the average number of children in families reported as a decimal value such as 2.37?

A Individual cases can only take on whole number values if a variable is discrete, but math can be done on these whole numbers to obtain meaningful fractional values.

An example should make it clearer how continuous numbers can take on in-between values. Weight is a continuous number. Suppose Tyrone stepped on a digital scale, one that weighs to the nearest pound, and reported, "I weigh 175 pounds." Does Tyrone weigh *exactly* 175 pounds? If there were a more precise scale, one that weighed to the nearest tenth of a pound, his weight might turn out to be 175.3 pounds. And, with an even more precise scale, one that measured to the nearest hundredth of a pound, his weight might now be found to be 175.34 pounds. How much Tyrone weighs depends on the precision of the scale used to measure him. Theoretically, continuous numbers can always be made more exact by using a better—more precise—measuring instrument.

Tyrone could weigh exactly 175 pounds, but he probably doesn't. Statisticians think of his weight as falling within an *interval* that has a midpoint of 175. The question is this: How wide is the interval? How much or how little does Tyrone really weigh?

The answer is that Tyrone weighs somewhere from 174.5 to 175.5 pounds. The scale reports weight to the nearest pound, so the unit of measurement is 1 pound. Tyrone's real weight falls somewhere in the interval from half a unit of measurement below his reported weight to half a unit of measurement above his reported weight. That's from $175 - 0.5 = 174.5$ to $175 + 0.5 = 175.5$.

If Tyrone weighed less than 174.5, say, 174.2, then the scale would have reported his weight as 174, not 175 pounds. And if he weighed more than 175.5, say, 175.9, it would have reported his weight as 176 pounds. It is not clear where in the range from 174.5 to 175.5 his weight really falls, but his weight does fall in that range. His weight—a continuous number—is reported as the midpoint (175) of a range of values, any of which could be his real weight.

A single continuous number represents a range of values. To help understand this, think back to high school chemistry and measuring how acidic or basic a liquid was. This was done by dipping pH paper into the liquid and comparing the color of the pH paper to a key. An example of a key for pH is shown in Figure 2.5.

Suppose one dipped a piece of the pH paper from Figure 2.5 into a solution and it turned the shade associated with a pH of 6. Is the pH of the solution exactly 6? The pH is closer to 6 than to 4, or else the paper would have turned the lighter shade for 4. It is also closer to 6 than to 8, because the paper was not the darker shade for 8. The pH of the solution is closer to 6 than to 4 or 8, but it probably is not exactly 6.

In what range does the pH of the solution fall? This pH paper measures to the nearest 2 points so that is the unit of measurement. Half of the unit of measurement is 1, so subtracting that from 6 and adding it to 6 give the range within which the pH of the solution really falls: between 5 and 7. When this pH paper says 6, it really means somewhere between 5 and 7.

Those pH values, 5 and 7, are called the real limits of the interval. **Real limits** are the upper and lower bounds of a single continuous number or of an interval in a grouped frequency distribution for a continuous variable.

Table 2.10, the grouped frequency distribution of IQ for the 68 sixth graders, can be used to examine the real limits for an interval in a grouped frequency distribution. Here, IQ is measured to the nearest whole number. However, it would be possible

Figure 2.5 Key for pH Paper pH is a continuous measure. If this pH paper turns the middle shade when dipped into a solution, the pH is called 6. But the pH of the solution may not be exactly 6. It is just closer to 6 than it is to 4 or 8.

to make a test that measures IQ to the nearest tenth or hundredth of a point, so IQ is a continuous variable. In Table 2.10, the bottom IQ interval ranges from 60 to 69, what are called the apparent limits for the interval. These are called **apparent limits** because they represent how wide the interval appears to be. But an IQ score of 60, at the bottom of the interval, really ranges from 59.5 to 60.5 and a score of 69, at the top of the interval, really ranges from 68.5 to 69.5. So, the scores in the 60 to 69 interval really fall somewhere in the range from 59.5, the real lower limit of the interval, to 69.5, the real upper limit. For continuous measures, real limits make researchers aware of how (im)precise their measures are.

Worked Example 2.2

The next section of this chapter turns to making graphs. Differentiating between discrete numbers and continuous numbers is important when figuring out what graph to make, so here's more practice.

I. A professor surveys a group of college students about how much time, in minutes, they spend on schoolwork per week. She makes a grouped frequency distribution that has intervals of 0–59 minutes, 60–119 minutes, 120–179 minutes, and so on. Is this a discrete measure or a continuous measure? What are the real limits for the 60–119 interval? For that interval, what are the interval width and the midpoint?

To figure this out, follow the flowchart in Figure 2.4. Minutes is a ratio-level variable, so it can be either discrete or continuous. To decide, one needs to think if the scale could be made more precise and to measure minutes to fractional values. The answer is yes. So, the amount of time spent studying is a continuous variable. The real lower limit for the 60–119 interval is 59.5, which is half a unit of measurement below the apparent lower limit of the interval. Similarly, the real upper limit for the interval is 119.5, half a unit of measurement above the apparent upper limit. The interval width is 60, the distance between the real limits of an interval, calculated by 119.5 – 59.5. And the midpoint of the interval is 89.5, halfway between 60 and 119, calculated by $\frac{60 + 119}{2}$.

II. A developmental psychologist finds out, from a national sample of teenagers, how many texts they send per week. He ends up making a grouped frequency distribution that has intervals of 1–25, 26–50, 51–75, and so forth. Is this a continuous or a discrete measure? For the second interval, 26–50, what are the real limits? For that interval, what are the interval width and the midpoint?

Again, use the flowchart in Figure 2.4 to figure out whether this is a continuous or a discrete measure. The number of texts sent is a ratio-level number, so it could be either continuous or discrete. A text is a text is a text, whether it is one character long or 160. One can't measure the number of texts more precisely than with whole numbers, which means that this variable, number of texts sent, is a discrete number. If a person sends 26 texts, he or she has sent 26 texts, not somewhere from 25.5 to 26.5 texts. This means that the real limits of the interval are the same as the apparent limits: 26–50. The interval width is 25 and the midpoint is 38, halfway between 26 and 50, calculated by way of $\frac{26 + 50}{2}$.

Practice Problems 2.2

Review Your Knowledge

2.08 How does a continuous number differ from a discrete number?

2.09 If there were five cases in the interval 45–49 of a grouped frequency distribution for a continuous variable and their original raw scores were unknown, what value(s) should be assigned to them?

Apply Your Knowledge

2.10 For each variable, determine if it is continuous or discrete.

a. The amount of snowfall, in inches, recorded during winter in a city

b. The number of days, in the past 30, on which a person has consumed any alcohol

c. A person's level of depression, as measured by the percentage of questions on a scale that was endorsed in the depressed direction

2.11 Below are sets of consecutive intervals from frequency distributions for continuous variables. For the second interval in each set, the one in bold, tell the real lower limit; the real upper limit; the interval width, *i;* and the interval midpoint, *m.*

a. 15–19; **20–24;** 25–29

b. 205–245; **255–295;** 305–345

c. 1.1–1.2; **1.3–1.4;** 1.5–1.6

d. 1,000–2,000; **3,000–4,000;** 5,000–6,000

2.3 Graphing Frequency Distributions

"A picture is worth a thousand words" is a common saying. It's time to put that saying into practice and explore visual displays of frequency distributions. Graphs, the subject of this section, can be used to display frequency information. With graphs, the information leaps out, whereas one has to work harder to find the information in a frequency distribution table.

This chapter covers three different graphs for showing frequency: (1) bar graphs, (2) histograms, and (3) frequency polygons. Choosing which graph to use depends on whether the numbers are discrete or continuous. If the numbers are discrete, use a bar graph. If the data are continuous, use a histogram or a frequency polygon.

Choosing which graph to use depends on whether the numbers are discrete or continuous. If the numbers are discrete, use a bar graph. If the data are continuous, use a histogram or a frequency polygon.

Decisions about which graph to use are often based on the level of measurement for the data—nominal, ordinal, interval, or ratio. Nominal- and ordinal-level data are discrete and should be graphed with a bar graph. Interval- and ratio-level data may be discrete or continuous. If the variable is discrete, use a bar graph. If the variable is continuous, use a histogram or frequency polygon. In actual practice, histograms and frequency polygons are often made for interval variables and ratio variables, whether the variable is continuous or not. In these instances, histograms or frequency polygons are used even if the data are discrete. Figure 2.6 is a flowchart summarizing the decision rules about how to choose the correct graph for a frequency distribution.

Bar Graphs

Bar graphs are used to demonstrate the frequency with which the different values of discrete variables occur. Sex, whether one is male or female, is a discrete variable. Table 2.11 shows a frequency distribution for the sex of students in a psychology class.

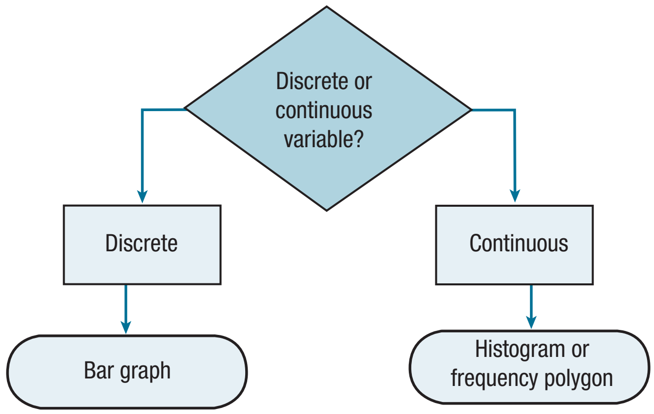


Figure 2.6 How to Choose: What Graph Should One Make for a Frequency Distribution? Start by using Figure 2.4 to determine whether the variable is continuous or discrete. The type of graph depends on this decision. No matter which graph one makes, be sure to give the graph a title and label all axes. If the variable being graphed is measured at the ordinal, interval, or ratio level, one can comment on the shape of the graph.

TABLE 2.11 Frequency Distribution of the Sex of 65 Students in an Upper-Level Psychology Class

Sex	Frequency (f)	Percentage (%)
Male	19	29.23
Female	46	70.77

Note: A frequency distribution for a nominal variable only provides frequency information and the percentage in each category. Because order is arbitrary for nominal variables, do not calculate cumulative frequencies or cumulative percentages.

Here's how to turn this table into a bar graph. Graphs are usually wider than tall, so start by making the X-axis longer than the Y-axis. **Figure 2.7** shows the template for the graph. The different categories of the variable, in this case male and female, go on the X-axis. The X-axis is labeled "Sex" and is marked "Male" and "Female."

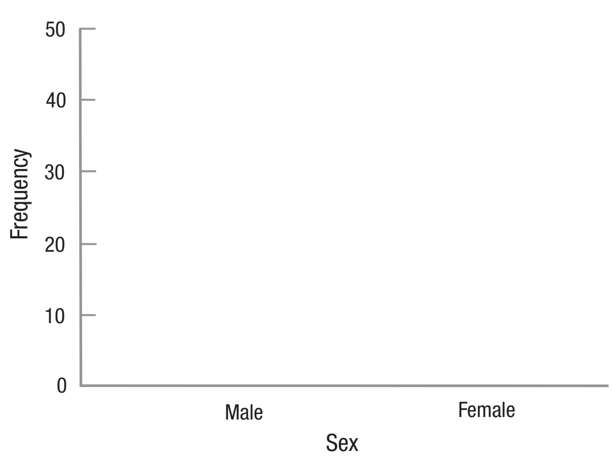


Figure 2.7 Template for Bar Graph Showing the Sex of Students ($N = 65$) in Upper-Level Psychology Class This "empty" graph shows how to set up a bar graph. Note that the graph is wider than tall, all axes are labeled, frequency goes on the Y-axis, and the categories being counted go on the X-axis.

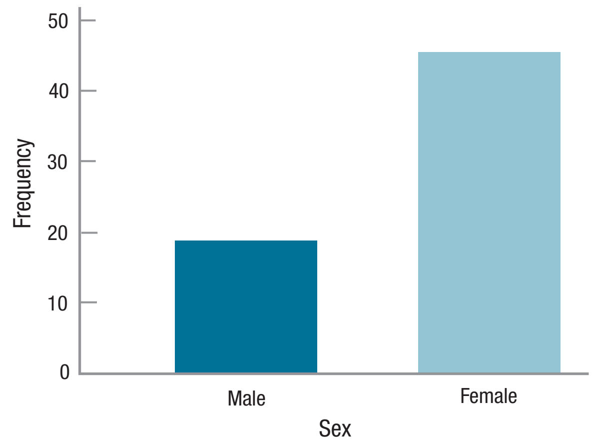


Figure 2.8 Bar Graph Showing the Sex of Students ($N = 65$) in Upper-Level Psychology Class Note that the bars don't touch each other in bar graphs because the variable on the *X*-axis is discrete.

Frequency goes on the *Y*-axis, so it is labeled "Frequency" and marked off in equal intervals, by 10s, from 0 to 50. Why go up to 50? Because the largest frequency was 46 and the axis has to accommodate that value.

Note that, just as with tables, the graph has a detailed title, *N* is mentioned, and everything is labeled clearly. Obviously, it is a good idea to use graph paper and a ruler.

To complete the bar graph, all one needs to do is draw a bar above each category on the *X*-axis. Note that the bar goes up as high as the frequency of the category in Figure 2.8, the completed figure. This is called a bar graph because of the bars; note that all of the bars are the same width and they don't touch each other. They don't touch each other because the variable is discrete.

The advantage of a picture over words, of the graph (Figure 2.8) over the table (Table 2.11), should be obvious. In this bar graph, it jumps out that there were a lot more women than men in this class.

Histograms

A **histogram** is a graphic display of a frequency distribution for continuous data. It is like a bar graph in that bars rise above the *X*-axis, with the height of the bars representing the frequencies in the intervals. However, a histogram differs from a bar graph in that the bars touch each other, representing the fact that the variable is continuous, not discrete. In Figure 2.9, a template for a histogram for the sixth-grade IQ data from earlier in the chapter is presented.

There are several things to note in this histogram:

- Even though the histogram is "empty" right now, it already has a detailed title.
- Both axes are clearly labeled.
- Frequency goes on the *Y*-axis and the values of the variable (in this case, IQ) go on the *X*-axis.
- The *Y*-axis starts at zero and is marked off by 5s to a height of 20. This accommodates the largest frequency in the data set, 15.

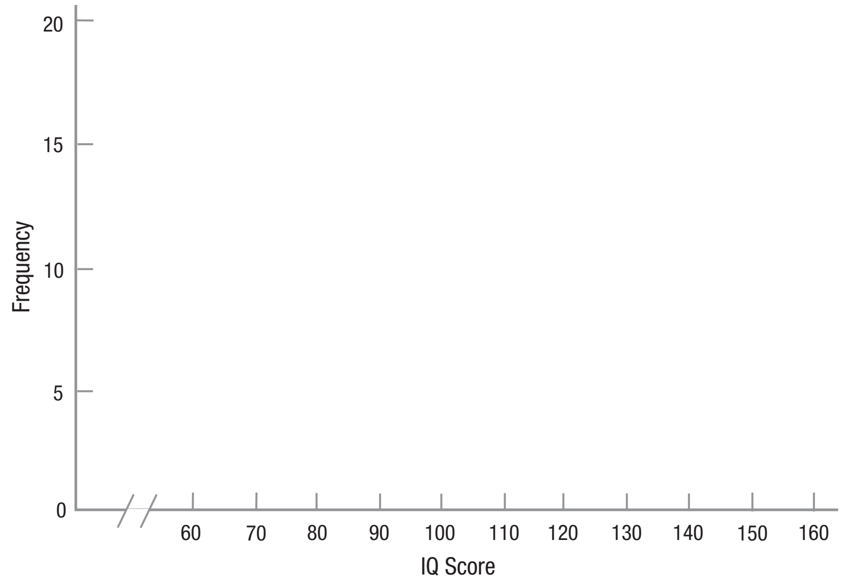


Figure 2.9 Template for Histogram Showing Grouped Frequency Distribution of IQ Scores for 68 Sixth Graders (Interval Width = 10) This "empty" graph shows how to set up a histogram. Note that frequency goes on the *Y*-axis, values of the variable being graphed on the *X*-axis, and that both axes are clearly labeled.

- The smallest value for IQ is 67, which is far away from a value of zero. If the *X*-axis started at zero, there would be a lot of blank space before the first interval. Instead, the *X*-axis starts at 60, the apparent lower limit of the first interval.
- Note the discontinuity mark on the *X*-axis, which alerts people viewing the graph that the axis doesn't start at zero.
- The highest IQ value is 152, so the *X*-axis ends at 160, what would be the start of the next higher interval.

Figure 2.10 shows the histogram once it has been completed. Note:

- The bars go up as high as the frequency in the interval, 15 for the highest frequency.
- Unlike a bar graph, the bars in a histogram touch each other because the variable being represented is continuous.
- There is one tricky thing in a histogram—the bars go up and come down at the real limits of the interval, not the apparent limits. The first interval, 60 to 69, for example, really covers scores from 59.5 to 69.5 and the bar reflects that. Look at the enlarged section of Figure 2.10 to get a clear picture of this.

Frequency Polygon

A **frequency polygon,** sometimes called a line graph, displays the frequency distribution of a continuous variable as a line, not with bars. A frequency polygon differs from a histogram in another way: the frequencies "go to zero" at the beginning (the far left) and the end (the far right) of the graph. Whether one uses a histogram or a frequency

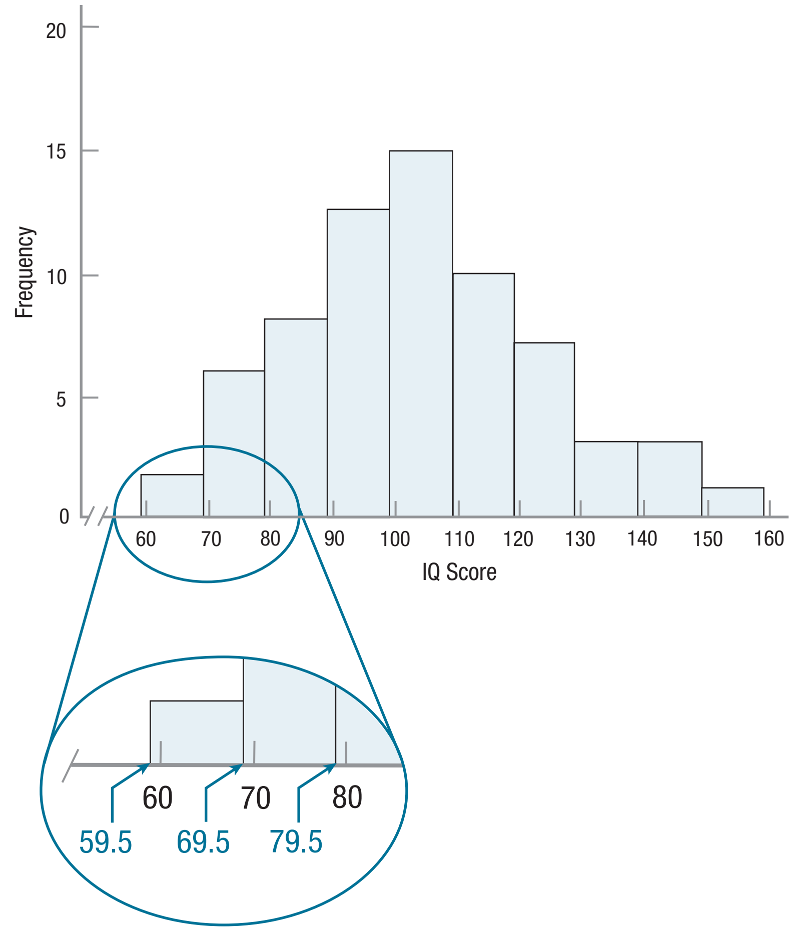


Figure 2.10 Histogram Showing Grouped Frequency Distribution of IQ Scores for 68 Sixth Graders (Interval Width = 10) Histograms graph the frequency of continuous variables, so the bars touch each other. Note that the bars are as wide as the real limits of each interval, not the apparent limits. This is highlighted in the enlarged section of the graph.

polygon to represent a frequency distribution is a matter of personal preference—they both are legitimate options for graphing frequencies for a continuous variable.

Figure 2.11 shows a template for a frequency polygon for the sixth-grade IQ data:

- Of course, there is a title and both axes are labeled.
- Frequency, on the *Y*-axis, starts at zero and goes up, in equal intervals, to the first interval above the largest possible frequency.
- The *X*-axis shows the midpoints of the intervals for the variable being graphed.
- The lowest IQ score in the data set is in the interval with a midpoint of 64.5 and the highest in an interval with a midpoint of 154.5. But, two “extra” midpoints are displayed on the *X*-axis, one (54.5) below the lowest interval and one (164.5) above the highest interval.

Figure 2.12 shows the completed frequency polygon. It was completed by placing a dot above each midpoint at the level of its frequency and then connecting the dots.

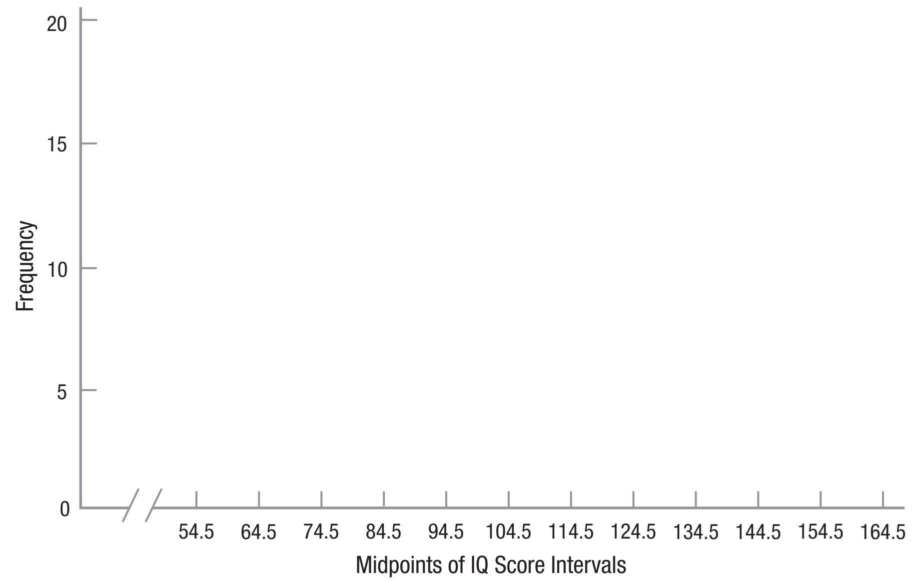


Figure 2.11 Template for Frequency Polygon Showing Grouped Frequency Distribution of IQ Scores for 68 Sixth Graders (Interval Width = 10) This "empty" graph shows how a frequency polygon should be set up. Frequency is marked on the *Y*-axis and midpoints of the IQ intervals on the *X*-axis. Both axes are clearly labeled and the graph is wider than tall.

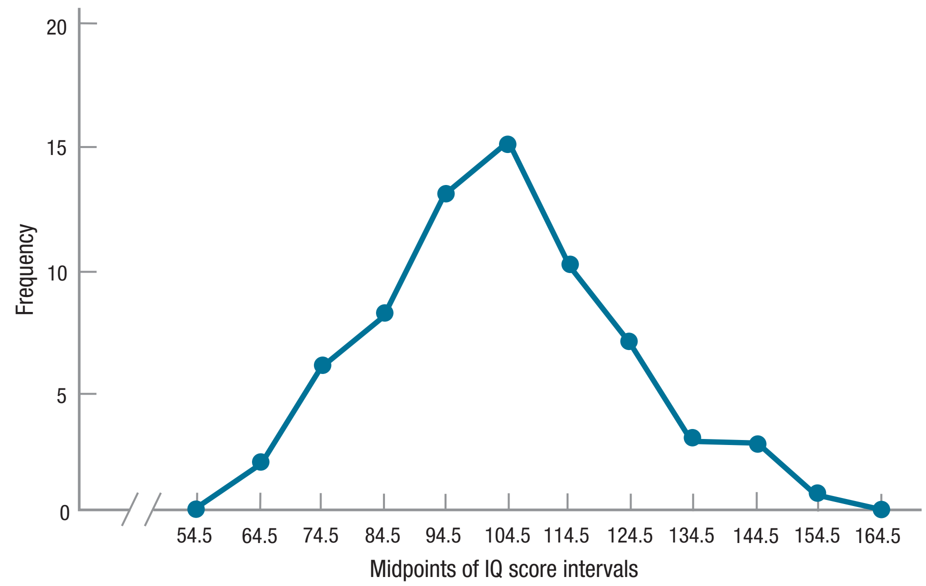


Figure 2.12 Frequency Polygon Showing Grouped Frequency Distribution of IQ Scores for 68 Sixth Graders (Interval Width = 10) Frequency polygons are made for continuous variables. The frequencies are marked by dots at the appropriate height at the midpoints of the intervals. Then the dots are connected by lines. Note that the graph starts and ends at intervals with a frequency of zero.

Note that the frequency for the very bottom and very top interval is zero. This is what was meant by saying that frequency polygons "go to zero" at the beginning and end of the graph.

There are a number of differences among the three graphs. Table 2.12 summarizes the differences among bar graphs, histograms, and frequency polygons.

TABLE 2.12 Key Differences Among Graphs

	Bar Graphs	Histograms	Frequency Polygons
Type of data	Discrete	Continuous	Continuous
Physical characteristics of the graph	✓ Bars don't touch each other.	✓ Bars touch each other. ✓ Bars go up and down at the real limits of the interval.	✓ Frequencies are marked with dots at the midpoints of the intervals. ✓ Dots are connected by lines. ✓ Frequencies "go to zero" at the far left and far right of the graph.

Worked Example 2.3

What type of graph could one make for the property crime rate frequency distribution in Table 2.5. The data are rates, so they are continuous and one could make either a histogram or a frequency polygon. This author is partial to histograms, but there would be nothing wrong with choosing to make a frequency polygon instead. The completed histogram—wider than tall, with crime rate on the *X*-axis and frequency on the *Y*, all axes labeled and with a clear title—is shown in Figure 2.13.

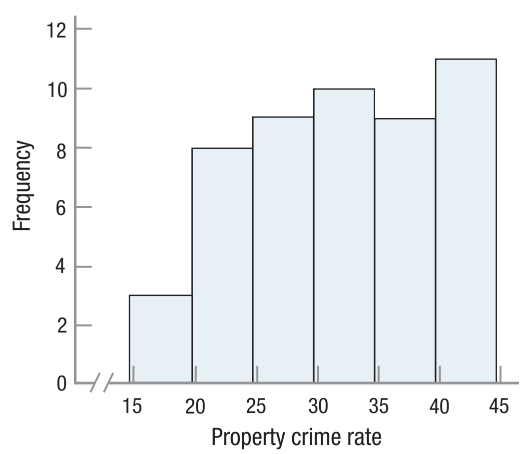


Figure 2.13 Histogram Showing Property Crime Rates per 1,000 Population for the 50 States This frequency distribution, which is for a continuous variable, could also be graphed as a frequency polygon.

Practice Problems 2.3

Review Your Knowledge

2.12 Match these two types of data, continuous and discrete, with the three types of graphs (bar graphs, histograms, and frequency polygons).

2.13 Where do the bars go up and come down for each interval on a histogram?

Apply Your Knowledge

2.14 A religion professor at a large university took a sample of 53 students and asked them what their religious faith was. Nine reported they were Muslim, 4 Buddhist, 2 Hindu, 16 Christian, 10 Jewish, 6 atheist, and 6 reported being other religions. Make a graph for these data.

2.15 A psychologist administered a continuous measure of depression to a representative sample of 772 residents of a midwestern state. Scores on the depression scale can range from 0 to 60 and higher scores indicate more depression. There were 423 with scores in the 0–9 range, 210 in the 10–19 range, 72 in the 20–29 range, 37 in the 30–39 range, 18 in the 40–49 range, and 12 in the 50–59 range. Make a graph for the data.

2.4 Shapes of Frequency Distributions

There are three aspects of the shape of frequency distributions to focus on—modality, skewness, and kurtosis.

Now that frequency distributions are being graphed, it is time to look at the shapes that data sets can take. There are three aspects of the shape of frequency distributions to focus on—modality, skewness, and kurtosis—and they are summarized in Table 2.13.

It is important to know the shape of a distribution of data. The shape will determine whether certain statistics can be used. In the next chapter, for instance, it will be shown that it's inappropriate to calculate a mean if a data set has certain irregular shapes.

TABLE 2.13 Three Aspects of Shape of Frequency Distribution Curves to Describe

Modality	Skewness	Kurtosis
How many high points there are in a data set	Whether the data set is symmetric	Whether a data set is peaked or flat

The shape of a data set only matters if the variable is measured at the ordinal, interval, or ratio level. With nominal variables, factors like race or religion, the order in which the categories are arranged is arbitrary. For example, Practice Problem 2.14 involved making a bar graph for the frequencies of different religions. The shape of this graph will vary, depending on whether one organizes the religions alphabetically, by ascending frequency, or by descending frequency. In contrast, one can't legitimately alter the order of the depression variable, graphed with either a histogram or frequency polygon, in Practice Problem 2.15.

One example of a common shape is Figure 2.14. Many call this a bell-shaped curve, but statisticians call it a normal curve or normal distribution. This curve has one highest point, in the middle, and the frequencies decrease symmetrically as the values move away from the midpoint. (Being symmetrical means that the left side of the curve is a mirror image of the right.) The frequency polygon of the sixth-grade IQ data in Figure 2.12 has a "normalish" shape.

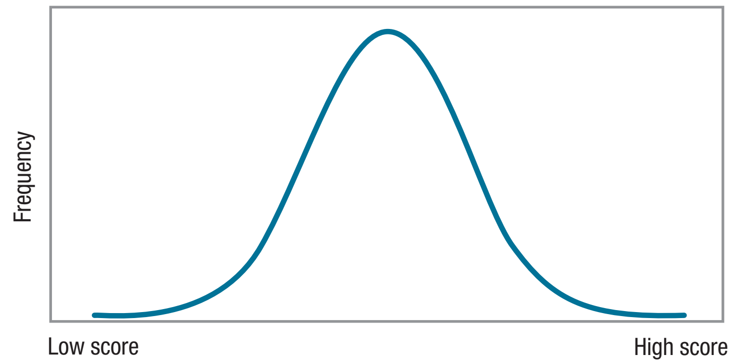


Figure 2.14 The Normal Distribution The normal distribution, sometimes called the bell-shaped curve, is highest at its midpoint and then shows symmetrical decreases in frequency as it moves to the tails.

Modality, the first of the three aspects used to describe shape, refers to how many peaks exist in the curve of the frequency distribution. A peak is a high point, also called a mode, and it represents the score or interval with the largest frequency. The normal curve has one peak, in the center of the distribution, and is called unimodal. A distribution can have two peaks, in which case it is called bimodal. If a distribution has three or more peaks, it is called multimodal. Figure 2.15 shows what distributions with different modalities look like.

Another characteristic of the shape of frequency distributions is **skewness,** which is a measure of how symmetric they are. A distribution is considered skewed if it is not symmetric. The normal curve is perfectly symmetric, as the left side and

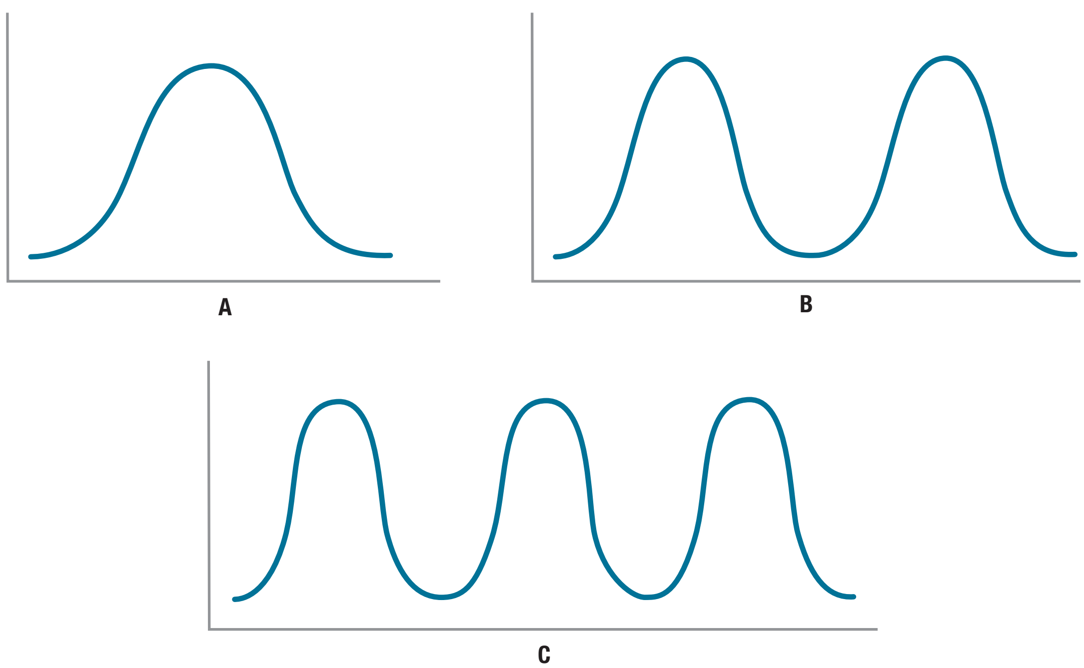


Figure 2.15 Examples of Modality Panel A shows a unimodal data set, panel B a bimodal data set, and panel C a multimodal data set.

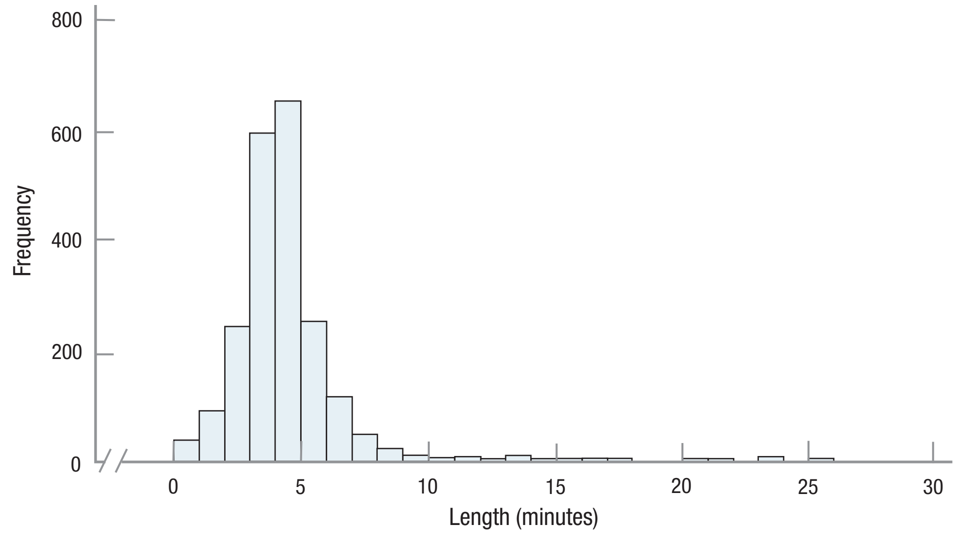


Figure 2.16 Length of Songs in an iTunes Library This histogram is a good example of positive skew, the data set not being symmetric but trailing off to the right. The "spike" around 4 minutes makes this a good example of a data set that is quite peaked. (Thanks to Samantha DeDionisio for collecting these data.)

right side are mirror images of each other. If there is asymmetry and the data tail off to the right, it is called **positive skew.** If the data tail off to the left, it is called **negative skew.** Positive and negative don't have good and bad connotations here. Statisticians simply use positive and negative to describe which side of the *X*-axis the tail is on, the right side for positive numbers and the left side for negative numbers.

For a good example of skewness, look at Figure 2.16. It shows the length of all 2,132 songs in a student's iTunes library. Most songs are around 3 or 4 minutes in length, but there are a few approximately 25 minutes long, giving this frequency distribution positive skew.

The third aspect of a frequency distribution is **kurtosis,** which is just a fancy term for how peaked or flat the distribution is. In Figure 2.14, the normal curve is neither too peaked nor too flat, so it has a normal level of kurtosis. Some sets of data have higher peaks, like the iTunes data in Figure 2.16. Other distributions are more flat, like the one shown in Figure 2.17 of the results from a single die rolled multiple times.

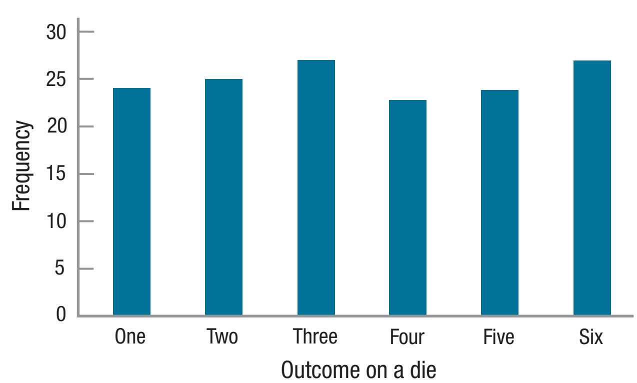


Figure 2.17 Frequency Distribution of Outcomes on a Die Rolled 150 Times Kurtosis refers to how peaked or flat a data set is. This data set is flat.

Stem-and-Leaf Displays

The final topic in this chapter, stem-and-leaf displays, is a wonderful way to summarize the whole chapter. A **stem-and-leaf display** is a combination of a table and a graph. It contains all the original data like an ungrouped frequency distribution table, summarizes them in intervals like a grouped frequency distribution, and "pictures" the data like a graph.

A stem-and-leaf display divides numbers into "stems" and "leaves." The leaves are the last digit on the right of a number. The stems are all the preceding digits. For the data in Table 2.4, the stem for South Dakota, North Dakota, and New Hampshire would be the tens digit 1, and the leaves would be, respectively, the ones digits 7, 9, and 9. With New York, the stem changes to the tens digit 2 and the leaf would be 0. All the numbers in Table 2.4 are two-digit numbers, ranging from numbers in the teens to numbers in the 40s, so the leaves are 1, 2, 3, and 4. Table 2.14 shows the first step in making a stem-and-leaf display, listing all the possible stems, followed by a vertical line.

TABLE 2.14 First Step in Constructing a Stem-and-Leaf Display

1	
2	
3	
4	

The first step in constructing a stem-and-leaf display is placing the stems.

The next step is to start tallying the leaves for each stem. When the leaves for all 50 states have been added, the results will look like Table 2.15. Note that the leaves for each stem are in order from low to high. Stem-and-leaf displays are a great way to organize data in preparation for making a grouped frequency distribution.

TABLE 2.15 Stem-and-Leaf Display Showing Property Crime Rates per 1,000 Population for the 50 United States

1	799
2	02234444555668899
3	0001224444555577789
4	00011111234

The three values in the top row are 17, 19, and 19.

Worked Example 2.4

For practice using a graph to figure out the shape of a data set, here are some data collected by a student, Neil Rufenacht. He took a ruler to a fast-food restaurant, bought several orders of french fries, and measured how long they were to the nearest tenth of a centimeter (cm). His results are shown as a stem-and-leaf display in Table 2.16 and as a histogram in Figure 2.18.

It is apparent in the histogram that the smallest french fries fell in the interval with a midpoint of 2 cm and the largest in the 14-cm interval. (That's from about three quarters of an inch to five and a half inches for those who prefer the English system of measurement.)

TABLE 2.16 Stem-and-Leaf Display for French Fry Data

1	79
2	5
3	345
4	000146667889
5	0012234445668
6	033333555668999
7	0001444567778889999
8	1234
9	00000235678
10	133444555
11	23466
12	0569
13	
14	2

This displays the same data as seen in the histogram in Figure 2.18. It has the same "normalish" shape. But, because the intervals are narrower, it shows more detail.

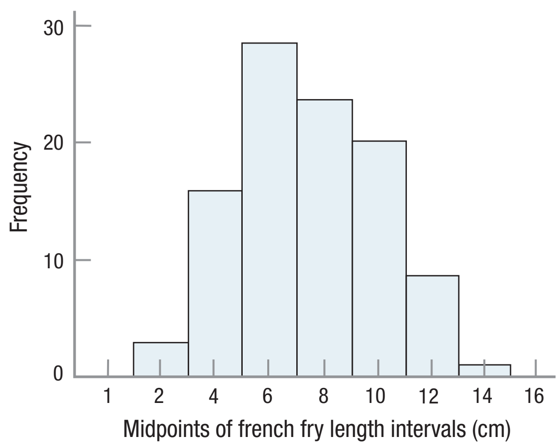


Figure 2.18 Histogram for Length of French Fries With one mode near the center and somewhat symmetrical tailing off of frequencies from the mode, this data set has a "normalish" shape.

The histogram has one mode, near the center, that doesn't seem to go up unusually high. This means that the distribution is unimodal and neither too peaked nor flat. In addition, the frequencies tail off, in a somewhat symmetrical fashion, as the lengths move away from the mode. The shape is not a perfect normal distribution, but it is "normalish." It seems reasonable to conclude that lengths of french fries, at least at this restaurant, resemble a normal distribution.

The stem-and-leaf display shows the same general shape, but offers more details. One can tell, for example, that the smallest french fry is 1.7 cm long.

DIY

Find something you can measure 100 times. Here are some ideas:

- Light a match and time how long it takes to burn down. Get another match, light it, and time the burning. Do this 98 more times.
- Find the prices of 100 stocks.
- Find the length, in seconds, of 100 songs from your iTunes library.
- Stick your tongue in a bowl of Cheerios and count how many stick to your tongue.
- Dip a tablespoon into a bowl of change and tally the value of the coins you picked up.
- Weigh, in grams, 100 eggs.

After you have collected your 100 data points, graph them and examine the shape. Given what you measured, does it make sense? For example, if you recorded the length of 100 songs from your music library and the distribution was positively skewed, does that, on reflection, seem reasonable?

Practice Problems 2.4

Review Your Knowledge

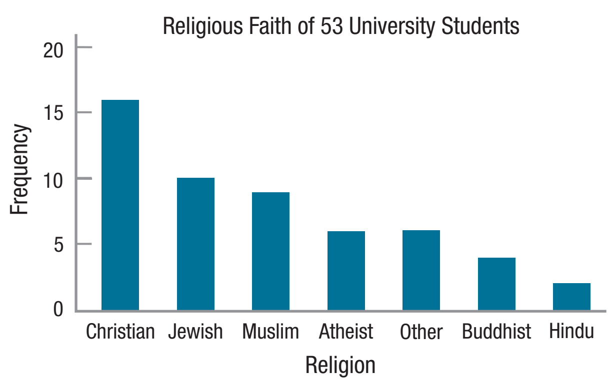
2.16 For what levels of measurement can one describe the shape of a frequency distribution?

2.17 Describe the normal curve in terms of symmetry, modality, and kurtosis.

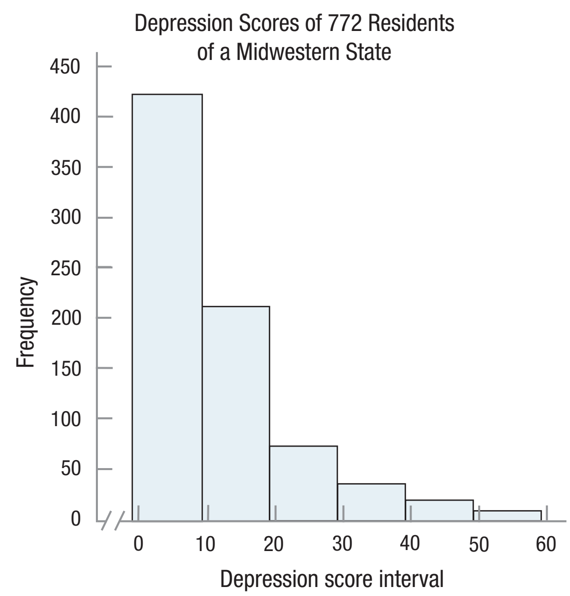
2.18 What does it mean if a frequency distribution is positively skewed?

Apply Your Knowledge

2.19 Here is a graph for the religious affiliation of the 53 students surveyed in Practice Problem 2.14. Determine the shape of the data set.



2.20 This graph shows the depression scores of the 772 residents surveyed in Practice Problem 2.15. Determine the shape of the data set.



2.21 Make a stem-and-leaf display and comment on the shape of the distribution for these data: 183, 192, 203, 203, 211, 219, 220, 227, 228, 228, 229, 230, 233, 234, 234, 248, 249, 254, and 266.

Application Demonstration

Data showing the suicide rates for all 50 U.S. states were obtained from the *Statistical Abstract of the United States*. The rates are expressed as the number of deaths per 100,000 population. The rates range from a low of 6.5 deaths per 100,000 population in New Jersey to a high of 21.7 per 100,000 population in Wyoming. To put that in perspective, 0.0065% of the population of New Jersey committed suicide in 2006 vs. a rate more than three times higher in Wyoming, 0.0217%.

The data are arranged in ascending order in Table 2.17 in preparation for making a frequency distribution. Note that the rates are reported to the nearest tenth, so that is the unit of measurement. There are a large number of unique values for the 50 states, but all values don't need to be maintained. Instead, those close to each other can be grouped together into intervals without losing vital information. Following the flowchart in Figure 2.3, a grouped frequency distribution makes sense.

TABLE 2.17 2006 U.S. Suicide Rate Data Arranged in Order

6.5	10.3	11.2	13.0	15.2
6.6	10.4	11.4	13.3	15.6
6.7	10.6	11.6	13.5	15.8
7.8	10.8	11.9	13.6	16.0
8.0	11.0	11.9	13.8	16.0
8.1	11.0	12.0	14.1	18.0
8.6	11.1	12.2	14.2	19.5
9.2	11.1	12.3	14.6	19.7
9.2	11.1	12.4	15.0	20.0
10.0	11.2	12.6	15.2	21.7

Note: Though the range of scores is not large, there are many unique values for these scores. A grouped frequency distribution makes more sense for these data than an ungrouped frequency distribution.

The next question is how many intervals to include. With a range of values from 6.5 to 21.7, the high score and low score are 15.2 points apart. An interval width of 2 points would mean having eight intervals. This fits in with the rule of thumb of including from five to nine intervals suggested in Table 2.5.

Intervals can't overlap. If the bottom interval starts at 6 and is 2 points wide, then the next interval starts at 8. Suicide rates are reported to the nearest tenth, so the apparent limits of the first interval are from 6.0 to 7.9. The bottom interval starts at 6. The real limits of the interval run from half a unit of measurement below the apparent bottom of the interval to half a unit of measurement above the apparent top of the interval. That is, from 5.95 to 7.95. *Note:* The distance between the real limits is the interval width.

The midpoints of the intervals are the points halfway between the interval's limits. The midpoint for the first interval, which ranges from 6.0 to 7.9, is calculated:

$$\frac{6.0 + 7.9}{2} = 6.95$$

The next midpoint is one interval width higher:

$$6.95 + 2.00 = 8.95$$

After calculating the other midpoints, one finds the frequencies by counting the number of cases in each interval. These frequencies are placed in a grouped frequency distribution table, as shown in Table 2.18.

TABLE 2.18 Partially Completed Grouped Frequency Distribution for Suicide Rate Data for the United States (Interval Width = 2.00)

Suicide Rate (Deaths per 100,000)	Interval Midpoint (*m*)	Frequency (*f*)	Cumulative Frequency (f_c)	Percentage (%)	Cumulative Percentage ($\%_c$)
20.0–21.9	20.95	2			
18.0–19.9	18.95	3			
16.0–17.9	16.95	2			
14.0–15.9	14.95	8			
12.0–13.9	12.95	10			
10.0–11.9	10.95	16			
8.0–9.9	8.95	5			
6.0–7.9	6.95	4			

Note: This partially completed frequency distribution contains the basic information.

Table 2.18 contains only the intervals, the midpoints, and the frequencies. Using Figure 2.1 as a guide, stair-step up to find the cumulative frequencies. Then use Equation 2.1 to transform frequencies into percentages and Equation 2.2 to transform cumulative frequencies into cumulative percentages in order to complete the grouped frequency distribution (see Table 2.19).

TABLE 2.19 Grouped Frequency Distribution for Suicide Rate Data for the United States (Interval Width = 2.00)

Suicide Rate (Deaths per 100,000)	Interval Midpoint (*m*)	Frequency (*f*)	Cumulative Frequency (f_c)	Percentage (%)	Cumulative Percentage ($\%_c$)
20.0–21.9	20.95	2	50	4.00	100.000
18.0–19.9	18.95	3	48	6.00	96.00
16.0–17.9	16.95	2	45	4.00	90.00
14.0–15.9	14.95	8	43	16.00	86.00
12.0–13.9	12.95	10	35	20.00	70.00
10.0–11.9	10.95	16	25	32.00	50.00
8.0–9.9	8.95	5	9	10.00	18.00
6.0–7.9	6.95	4	4	8.00	8.0

Note: This table summarizes the suicide rate, at a state level, in the United States.

The next step is to make a graph of the data. Use Figure 2.4 to determine whether the data are continuous and Figure 2.6 to determine whether a histogram or frequency polygon should be made. Figure 2.19 displays a histogram.

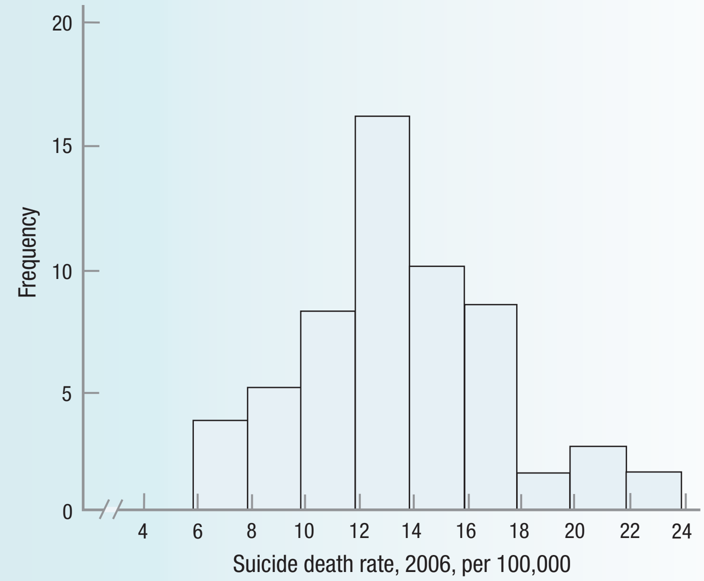


Figure 2.19 Histogram for U.S. Suicide Rates This histogram, with an interval width of 2.00, is unimodal and seems a little more peaked than normal. It might have a little positive skew.

This is a histogram, so the bars go up and come down at the real limits of the intervals, not the apparent limits. It may be hard to see, but the first bar goes up at 5.95, not 6.00, and comes down at 7.95, not 8.00.

DIY

Every year the U.S. Census Bureau used to publish a compendium of odd facts about America and Americans. Want to know the dollar value of all the crops and livestock produced in a state? The *Statistical Abstract of the United States* will tell you that it ranges from $32 million in Alaska to $35 billion in California. California also leads in the number of federal and state prisoners (171,000), while Hawaii has the longest life expectancy, 77 years.

Though the last *Statistical Abstract* was published in 2012, it is still available on the Web and in libraries. Make a reference librarian happy and ask to see a copy. Or Google the term and explore it online. Whatever approach you take, get hold of a copy, find a table that has data of interest to you, and reduce that data into a frequency distribution and graph. By doing so, what did you learn about your variable?

SUMMARY

Make a frequency distribution for a set of data.

- Frequency distributions summarize a set of data by tallying how often the values, or ranges of values, occur. For all but nominal-level data, they may also display information about cumulative frequency, percentage, and cumulative percentage.

Decide if a number is discrete or continuous.

- Discrete numbers are whole numbers that answer the question “How many?” Continuous numbers can be fractional and answer the question “How much?” The number of decimal places reported for a continuous variable depends on the precision of the measuring instrument.

- The range within which a single continuous value, or an interval of continuous numbers, falls is bounded by the real limits of the interval.

Choose and make the appropriate graph for a frequency distribution.

- Frequency distributions of discrete data are graphed with bar graphs; distributions of continuous data are graphed with histograms or frequency polygons.

Describe modality, skewness, and kurtosis for a frequency distribution.

- Modality refers to how many high points there are in a data set. If a distribution is asymmetric, it is skewed, while kurtosis refers to how peaked or flat a distribution is.

Make a stem-and-leaf display.

- Stem-and-leaf displays are a great way to summarize a set of data. They present the data compactly, like a grouped frequency distribution; keep all the details like an ungrouped frequency distribution; and show the shape, like a graph.

KEY TERMS

apparent limits – what seem to be the upper and lower bounds of an interval in a grouped frequency distribution.

bar graph – a graph of a frequency distribution for discrete data that uses the heights of bars to indicate frequency; the bars do not touch.

continuous numbers – answer the question "how much" and can have "in-between" values; the specificity of the number, the number of decimal places reported, depends on the precision of the measuring instrument.

cumulative frequency – a count of how often a given value, or a lower value, occurs in a set of data.

cumulative percentage – cumulative frequency expressed as a percentage of the number of cases in the data set.

discrete numbers – answer the question "how many," take whole number values, and have no "in-between" values.

frequency distribution – a tally of how often different values of a variable occur in a set of data.

frequency polygon – a frequency distribution for continuous data, displayed in graphical format, using a line connecting dots, above interval midpoints, that indicate frequency.

grouped frequency distribution – a count of how often the values of a variable, grouped into intervals, occur in a set of data.

histogram – a frequency distribution for continuous data, displayed in graph form, using the heights of bars to indicate frequency; the bars touch each other.

kurtosis – how peaked or flat a frequency distribution is.

midpoint – the middle of an interval in a grouped frequency distribution.

modality – the number of peaks that exist in a frequency distribution.

negative skew – an asymmetrical frequency distribution in which the tail extends to the left, in the direction of lower scores.

positive skew – an asymmetrical frequency distribution in which the tail extends to the right, in the direction of higher scores.

real limits – what are really the upper and lower bounds of a single continuous number or of an interval in a grouped frequency distribution.

stem-and-leaf display – a data summary technique that combines features of a table and a graph.

skewness – deviation from symmetry in a frequency distribution, which means the left and right sides of the distribution are not mirror images of each other.

ungrouped frequency distribution – a count of how often each individual value of a variable occurs in a set of data.

CHAPTER EXERCISES

Answers to the odd-numbered exercises appear at the back of the book.

Review Your Knowledge

2.01 A frequency distribution is a ____ of how often the values of a variable occur in a data set.

2.02 There are two types of frequency distribution tables: ____ and ____.

2.03 A basic frequency distribution just contains information about what ____ occur in a data set and what their ____ are.

2.04 ____ tells how many cases in a data set have a given value or a lower one.

2.05 The abbreviation for cumulative frequency is____.

2.06 The cumulative frequency for the top row in a frequency distribution is equal to ____.

2.07 If one has ____ -level data in a frequency distribution, it should be organized in some logical fashion.

2.08 The ____ column in a frequency distribution is a restatement of the information in the cumulative frequency column.

2.09 When dealing with a variable that has a large range, a ____ frequency distribution usually makes more sense than an ungrouped frequency distribution.

2.10 As a general rule of thumb, grouped frequency distributions have from ____ to ____ intervals.

2.11 Having a small number of intervals may show the big picture, but the danger exists of losing sight of the ____ in the data set.

2.12 All intervals in a grouped frequency distribution should be the same ____, so the frequencies in different intervals can be compared meaningfully.

2.13 A ____ is the middle point of an interval.

2.14 If the value for a case in an interval is unknown, assign it the value associated with the ____.

2.15 Discrete numbers answer the question ____.

2.16 Discrete numbers take on ____ number values only.

2.17 ____- and ____-level numbers are discrete numbers.

2.18 ____ numbers answer the question "How much?"

2.19 Continuous numbers can have ____ values.

2.20 The specificity of a ____ depends on the precision of the instrument used to measure it.

2.21 Continuous numbers are always ____- or ____-level numbers.

2.22 Interval- and ratio-level numbers can be ____ or ____.

2.23 A single value of a continuous number really represents a ____ of values.

2.24 The ____ are the upper and lower bounds of a single continuous number or of an interval in a grouped frequency distribution for continuous numbers.

2.25 The distance between the real limits of an interval is ____.

2.26 Which graph is used to display a frequency distribution depends on whether the numbers in the frequency distribution are ____ or ____.

2.27 The graph used for a frequency distribution of discrete data is a ____.

2.28 To make a graph of a frequency distribution of continuous data, one can use a ____ or a ____.

2.29 In terms of size, graphs are usually ____ than ____.

2.30 Graphs should have a descriptive title and the ____ should be clearly labeled.

2.31 Frequency is marked on the ____-axis of the graph of a frequency distribution.

2.32 The intervals for the values of the variable being graphed are marked on the ____ -axis of a histogram.

2.33 In a histogram, the data are continuous, so the bars ____ each other.

2.34 In a frequency polygon, frequency is marked with a dot placed at the appropriate height above the ____of an interval.

2.35 One shouldn't look at the shape of a graph for ____-level data.

2.36 The ____ is symmetric, has the highest point in the middle, and includes frequencies that decrease as one moves away from the midpoint.

2.37 The three aspects of the shape of a frequency distribution that are described are ____, ____, and ____.

2.38 ____ describes how many peaks exist in a frequency distribution.

2.39 If a frequency distribution has two peaks, it is called ____.

2.40 A frequency distribution that tails off on one side is said to be ____.

2.41 If a frequency distribution tails off on the left-hand side, it has ____ skewness.

2.42 Kurtosis refers to how ____ or ____ a frequency distribution is.

2.43 A stem-and-leaf display summarizes a data set like a ___________ frequency distribution.

2.44 In a stem-and-leaf display for 21, 21, 24, 25, 28, 30, 31, 34, and 39, the numbers 2 and 3 would be the _____s and 0, 1, 4, 5, 8, and 9 would be the ________s.

Apply Your Knowledge

Making frequency distribution tables

2.45 A psychologist completes a survey in which residents in a psychiatric hospital are given diagnostic interviews in order to determine their psychiatric diagnoses. After each interview, the researcher counts how many diagnoses each resident has. Given the following data, make an ungrouped frequency distribution showing frequency, cumulative frequency, percentage, and cumulative percentage: 2, 3, 2, 1, 0, 1, 1, 4, 2, 3, 1, 0, 1, 2, 4, 3, 2, 2, 2, 2, 1, 2, 3, 2, 5, 2, 3, 2, 1, 1, 2, 3, 3, 3, 2, 2, 1, 1, 2, 2, 3, 1, 1, 2, and 2.

2.46 A cognitive psychologist wanted to investigate the stage of moral development reached by college students. She believed that people progressed, in order, through six stages of moral reasoning, from Stage I to Stage VI. She obtained a representative sample of college students in the United States and administered an inventory to classify stage of moral development. Here are the numbers of people classified at the different stages: I = 17, II = 34, III = 78, IV = 187, V = 112, and VI = 88. Make a frequency distribution table for these data, showing frequency, cumulative frequency, percentage, and cumulative percentage.

2.47 A college registrar completes a survey of classrooms on campus in order to find out how many usable seats there are in each one. Make a grouped frequency distribution for her data, using an interval width of 20 and an apparent lower limit of 10 for the bottom interval. Report midpoint, frequency, cumulative frequency, percentage, and cumulative percentage. Here are the data the registrar collected: 12, 26, 18, 17, 102, 20, 35, 46, 50, 28, 29, 53, 75, 30, 37, 45, 58, 43, 42, 36, 50, 60, 55, 45, 40, 23, 28, 38, 39, 40, 50, 60, 45, 36, 28, 40, 54, 62, 38, 58, and 24.

2.48 A professor of education was curious about students' expectations of academic success. She obtained a sample of college students at the start of their college careers and asked them what they thought their GPAs would be for the first semester. Below are the data she obtained. Make a grouped frequency distribution for the data, using an interval width of 0.5 and an apparent lower limit of 2.1 for the bottom interval. Report midpoint, frequency, cumulative frequency, percentage, and cumulative percentage.

3.9, 4.0, 3.0, 3.5, 3.0, 3.5, 3.5, 3.9, 3.5, 2.5, 3.5, 3.8, 3.9, 3.0, 3.9, 3.5, 3.9, 3.0, 3.5, 2.8, 2.4, 2.7, 2.6, 3.4, 3.6, 2.5, 3.8, 2.9, 2.6, 2.4

Discrete vs. continuous numbers

2.49 For each variable, decide if it is discrete or continuous:

a. The number of cases of flu diagnosed at a college in a given semester
b. The weight, in grams, of a rat's hypothalamus
c. The number of teachers, including substitutes, in a school district who are certified to teach high school math
d. The number of times that a rat pushes a lever in a Skinner box during a 15-minute period
e. How long, in seconds, it takes applause to die down at the end of a school assembly

2.50 For each variable, decide if it is continuous or discrete:
a. The depth, in inches, a person can drive a 3-inch nail with one hammer blow
b. The number of students in a classroom who are absent for at least one day during the month of February
c. The total number of days that students in a classroom are absent during the month of February
d. How many friends one has on Facebook
e. The amount of time, measured in seconds, that a person in a driving simulator keeps his or her eyes on the road while engaging in a cell phone conversation

Real vs. apparent limits

2.51 Below are sets of consecutive numbers, or intervals, from frequency distributions for continuous variables. For the second number (or interval) in each set, the one in bold, identify the real lower limit, real upper limit, interval width, and midpoint.
a. 27–30; **31–34;** 35–38
b. 10–20; **30–40;** 50–60
c. 2.00–2.49; **2.50–2.99;** 3.00–3.49
d. 10; **11;** 12

2.52 For each scenario, tell what the real limits of the number/interval are:
a. If a person's temperature is reported as 100.3 (and not 100.2 or 100.4), what is the interval within which the person's temperature really falls?
b. If a family has 7 children, not 6 or 8, what is the interval within which the number of children they have really falls?
c. If a person's IQ is reported as 115 (and not 114 or 116), what is the interval within which the person's IQ really falls?
d. If nations' populations are reported to the nearest million and the United States has a population of 321 million, not 320 or 322, how many people really live in the United States?
e. If a man has 3 televisions in his house, not 4 or 5, what is the interval within which the number of televisions he has really falls?

Graphs and shapes of distributions

2.53 A developmental psychologist observes how children interact with their parents and classifies the children's attachment as secure, avoidant, resistant, or disorganized. He classifies 33 children as securely attached, 17 as avoidant, 13 as resistant, and 21 as disorganized. Graph this frequency distribution and comment on its shape.

2.54 A music teacher held an assembly at an elementary school during which he explained the different types of instruments to the children. Afterwards, he surveyed them as to what kind of instrument each would like to play. Here are the results: percussion = 12, brass = 39, wind = 42, and string = 28. Graph the results and comment on the shape of the graph.

2.55 A psychologist administered an aggression scale to a sample of high school girls. Scores on the scale can range from 20 to 80. Fifty is considered an average score; higher scores mean more aggression. Twelve girls had scores in the 20–29 range; 33 were 30–39; 57 were 40–49; 62 were 50–59; 19 were 60–69; and 8 had scores in the 70–79 range. Graph the results and comment on the shape of the graph.

2.56 A researcher from a search engine company runs Internet searches and times how long each one takes. She found that 17 searches took 0.01 to 0.05 seconds; 57 took 0.06 to 0.10 seconds; 134 took 0.11 to 0.15 seconds; 146 took 0.16 to 0.20 seconds; 398 took 0.21 to 0.25 seconds; 82 took 0.26 to 0.30 seconds;

and 56 took 0.31 to 0.35 seconds. Graph the frequency distribution and comment on its shape.

Making stem-and-leaf displays

2.57 Make a stem-and-leaf display for these data and comment on the shape of the distribution: 11, 12, 21, 23, 27, 27, 29, 30, 30, 33, 34, 35, 39, 43, 45, 47, 53, 53, 67, 75, 84, and 96.

2.58 Make a stem-and-leaf display for these data and comment on the shape of the distribution: 8, 9, 13, 13, 16, 21, 22, 24, 25, 25, 29, 33, 33, 45, 48, 55, 58, 61, 63, 64, 66, 66, 69, 71, 83, 92, 93, and 95.

Expand Your Knowledge

2.59 If the real limits and apparent limits for an interval are the same, then the variable is:

- **a.** continuous.
- **b.** discrete.
- **c.** interval.
- **d.** ratio.
- **e.** Real and apparent limits are never the same.

2.60 For a row in the middle of a frequency distribution, either ungrouped or grouped, which statement is probably true?

- **a.** $f > f_c$
- **b.** $f_c > f$
- **c.** $f = f_c$
- **d.** None of these statements is ever true.

2.61 If a researcher has a grouped frequency distribution for a continuous variable and there are five people in the interval ranging from 50 to 54, what X values should he assign them if he wants to assign values?

- **a.** He should randomly select values from 50 to 54.
- **b.** He should randomly select values from 49.5 to 54.5.
- **c.** As there are five people and there are five integers from 50 through 54, he should assign them values of 50, 51, 52, 53, and 54.
- **d.** 50
- **e.** 52
- **f.** 52.5
- **g.** 54

2.62 The distribution of prices of all the new cars sold in a year, including Ferraris, Lamborghinis, and other high-end cars, is probably:

- **a.** normally distributed.
- **b.** positively skewed.
- **c.** negatively skewed.
- **d.** not skewed.

2.63 Diane wears a digital watch that reports the time to the nearest minute. At 10:16 A.M., by her watch, she left her office. She returned at 10:21 A.M. To the nearest minute, what is the shortest amount of time that she was out of her office? What is the longest?

2.64 Make a stem-and-leaf display for these data. Use an interval width of 5.
120, 122, 123, 123, 124, 125, 126, 125, 126, 130, 134, 135, 135, 138, 147

SPSS

SPSS can be used to generate frequency distribution tables and graphs like those created in this chapter. The "Frequencies" command is used to generate an ungrouped frequency distribution. Figure 2.20 shows how to access the Frequencies command by clicking on "Analyze" and then "Descriptive Statistics."

Once the menu shown in Figure 2.21 has been opened, use the arrow key to shift the variable one wants to use from the box on the left to the "Variable(s)" box. Here, the variable name is "Children."

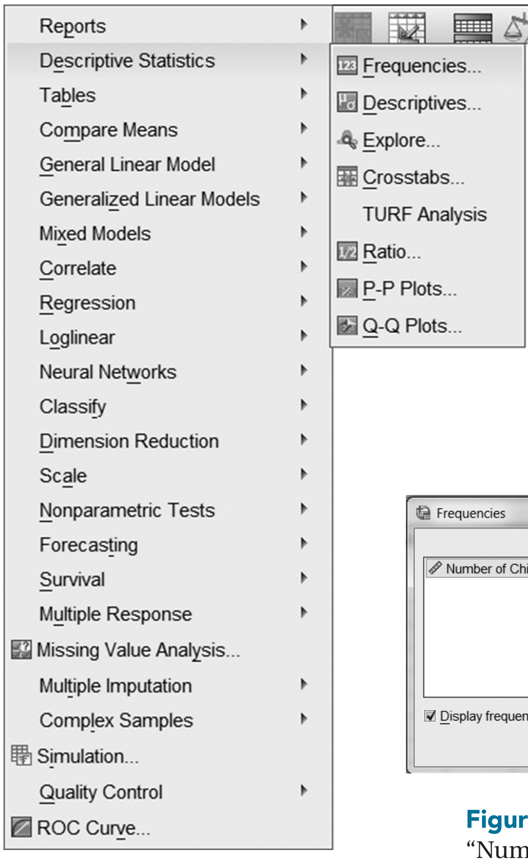


Figure 2.20 Starting the Frequencies Command The "Frequencies" command is accessed under "Analyze" and "Descriptive Statistics." (Source: SPSS)

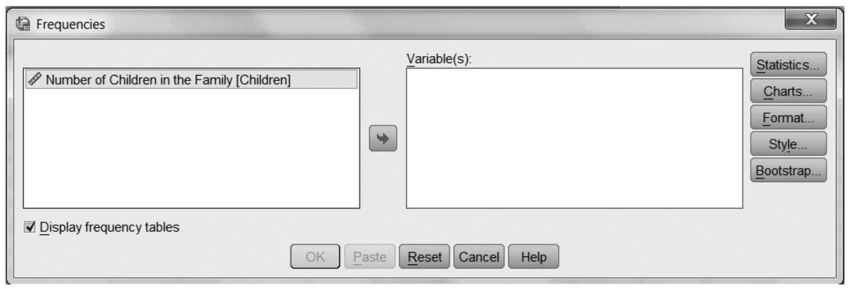


Figure 2.21 The Frequencies Command Menu Once a variable like "Number of children" has been moved into the "Variable(s)" box, the "OK" button on the lower right becomes active. (Source: SPSS)

Once there's a variable in the Variable(s) box, click on "OK" to generate an ungrouped frequency distribution. The Frequencies output can be seen in Figure 2.22.

Making a grouped frequency distribution in SPSS is more involved. One has to use "Transform" commands like "Recode" or "Visual binning" to place the data into intervals before using Frequencies.

Graphing is easy with SPSS. To start making a graph, click on "Graphs" as shown in Figure 2.23.

Number of Children in the Family

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	9	29.0	29.0	29.0
	2.00	14	45.2	45.2	74.2
	3.00	5	16.1	16.1	90.3
	4.00	2	6.5	6.5	96.8
	6.00	1	3.2	3.2	100.0
	Total	31	100.0	100.0	

Figure 2.22 Example of Frequencies Output from SPSS Compare this SPSS output to Table 2.3, for the same data. Note that Table 2.3 was arranged upside down, compared to SPSS. Table 2.3 also contained additional information—cumulative frequency. Finally, Table 2.3 included values, such as five children in the family, for which the frequency was zero. (Source: SPSS)

Clicking on “Chart Builder. . .” opens up the screen shown in Figure 2.24. Note that the list of graphs SPSS can make may be seen on the bottom left.

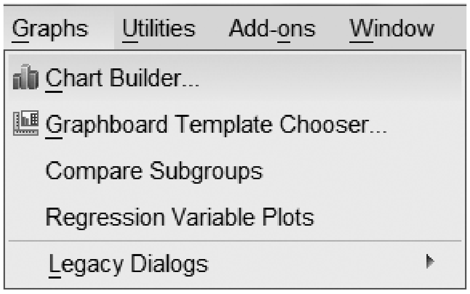


Figure 2.23 Starting the Graph Command in SPSS To start a graph, click on “Graphs” and then on “Chart Builder. . . .” (Source: SPSS)

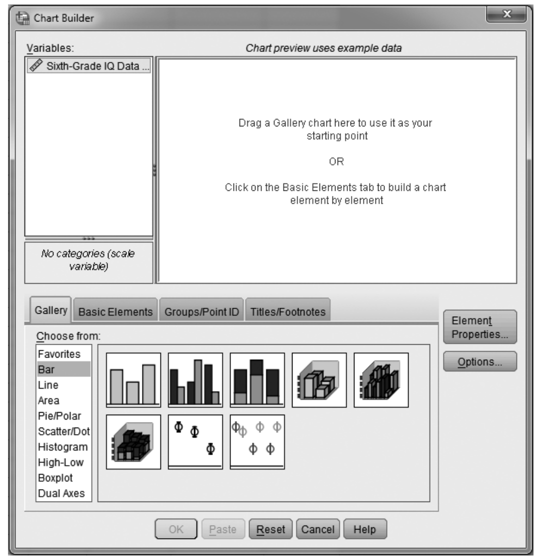


Figure 2.24 Chart Builder Screenshot The list on the bottom left shows the different categories of graphs available, while the pictures to the right show the different options within a category. In the list, there are bar graphs and histograms; look under “Line” to find frequency polygons. (Source: SPSS)

Figure 2.25 shows how to make a histogram. Histogram was highlighted in the list of graphs and the mouse was used to drag the basic histogram from the gallery of histograms to the preview pane.

To finish the histogram commands, see Figure 2.26. Note that the variable for which the frequencies are being graphed, IQ, has been highlighted and dragged to the *X*-axis.

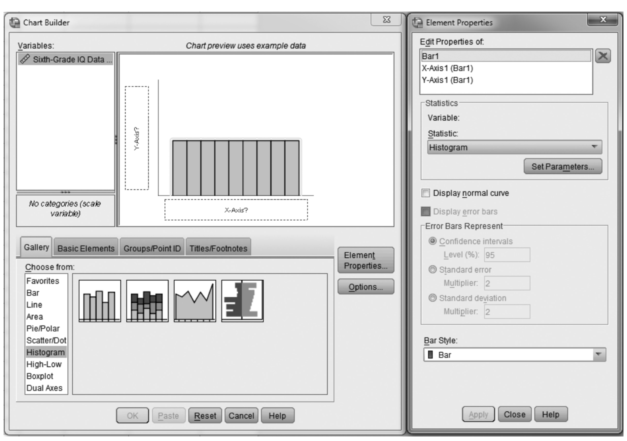


Figure 2.25 SPSS Commands for Making a Histogram Histogram was highlighted in the gallery. Then the basic histogram was clicked and dragged into the preview pane. (Source: SPSS)

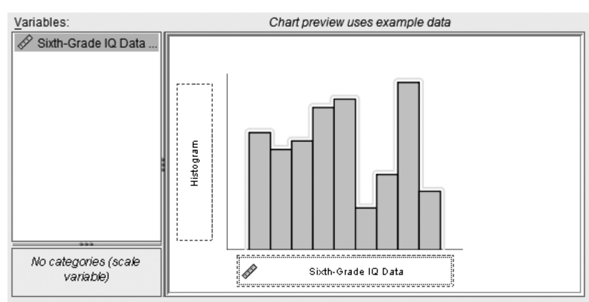


Figure 2.26 Finishing the Histogram Commands As the final step, highlight the variable for which the histogram is being completed (here, the IQ score) and drag it to the *X*-axis. (Source: SPSS)

Once the "OK" button has been clicked, SPSS generates a histogram as seen in Figure 2.27. This is called the "default" version because SPSS decides how wide the intervals should be.

A major advantage of using SPSS to make a graph over making it by hand is that it is easy to edit a graph in SPSS. Double-click on the graph to open up the editor. With just a few clicks of the mouse, one can change the interval width, the scale on the *Y*-axis, the labels on the axes, or the size of the graph. Figure 2.28 shows an edited version of a histogram for the IQ data.

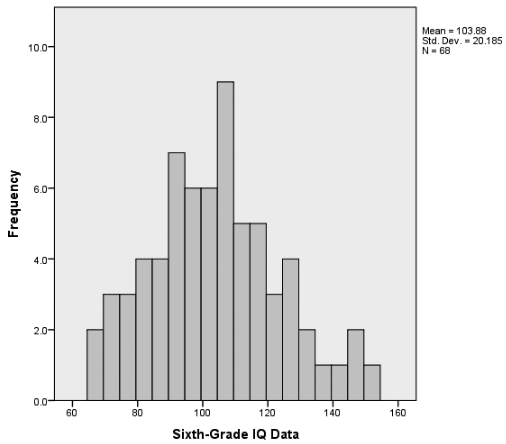


Figure 2.27 Histogram Generated by SPSS This is the histogram generated by SPSS when the "OK" button was pressed. (Source: SPSS)

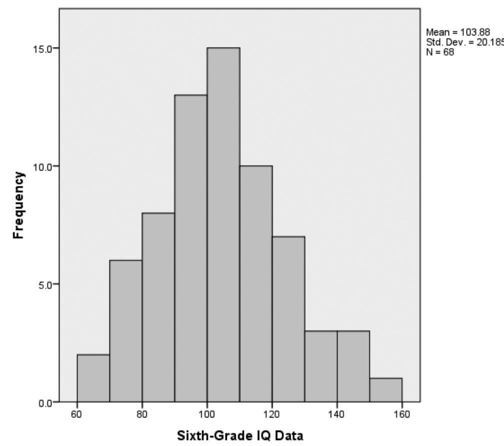


Figure 2.28 Edited Version of Histogram Editing graphs is easy in SPSS. Compare this version to the default graph shown in Figure 2.27. (Source: SPSS)

Stem-and-leaf plots can also be produced easily using SPSS. From the Analyze menu, choose "Descriptive Statistics" and then the "Explore" command as shown in Figure 2.29.

As shown in Figure 2.30, move the variable you want to analyze into the "Dependent List" by highlighting the variable and the clicking the arrow. Next,

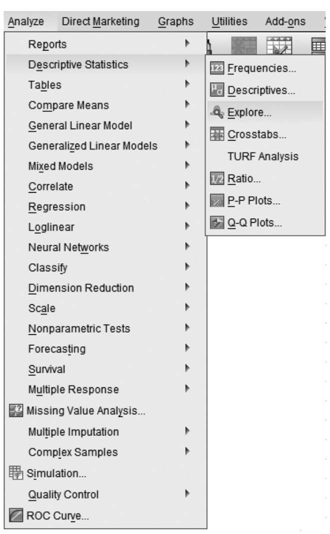


Figure 2.29 Starting Stem-and-Leaf Plot Commands Stem-and-leaf plots are accessed through the "Explore" command. (Source: SPSS)

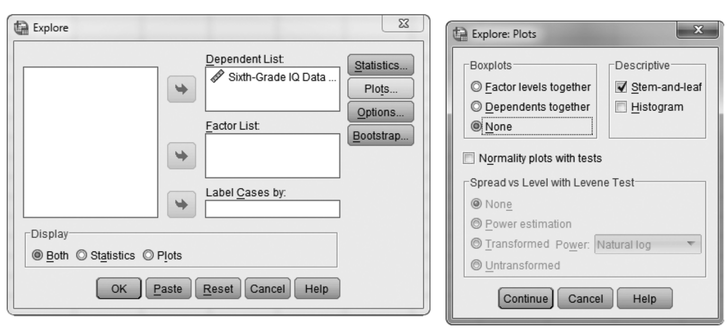


Figure 2.30 Commands for Making a Stem-and-Leaf Plot The variable to be plotted has been moved to "Dependent List" and "Stem-and-leaf" has been chosen. (Source: SPSS)

```
Sixth-Grade IQ Data Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2.00        6 .  78
     6.00        7 .  234569
     8.00        8 .  00245589
    13.00        9 .  0002344557788
    15.00       10 .  011223555678999
    10.00       11 .  1111256899
     7.00       12 .  1226889
     3.00       13 .  446
     3.00       14 .  278
     1.00       15 .  2

 Stem width:        10
 Each leaf:      1 case(s)
```

Figure 2.31 SPSS Output of a Stem-and-Leaf Display SPSS provides an output that includes the stem and leaf values as well as the frequencies. (Source: SPSS)

click "Plots" and make sure that under the Descriptive option "Stem-and-leaf" is checked. This is the default. You can change the default option under "Boxplots" to "None" because we are not running a boxplot.

After clicking "Continue" and "OK," the output in Figure 2.31 will be produced.