

## Histograms, Frequency Polygons, and Ogives

**Objective 2.** Represent data in frequency distributions using histograms, frequency polygons, and ogives.

After the data have been organized into a frequency distribution, they can be presented in graphic forms. The purpose of graphs in statistics is to convey the data to the viewer in pictorial form. It is easier for most people to comprehend the meaning of data presented graphically than data presented numerically in tables or frequency distributions. This is especially true if they have little or no statistical knowledge.

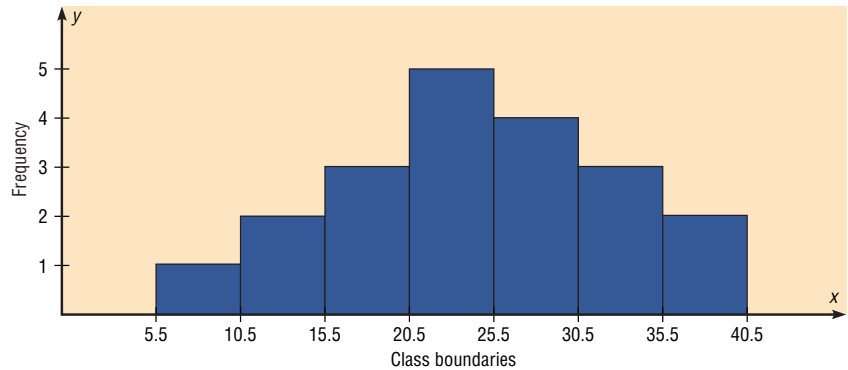
Statistical graphs can be used to describe the data set or analyze it. Graphs are also useful in getting the audience's attention in a publication or a speaking presentation. They can be used to discuss an issue, reinforce a critical point, or summarize a data set. They can also be used to discover a trend or pattern in a situation over a period of time.

The three most commonly used graphs in research are

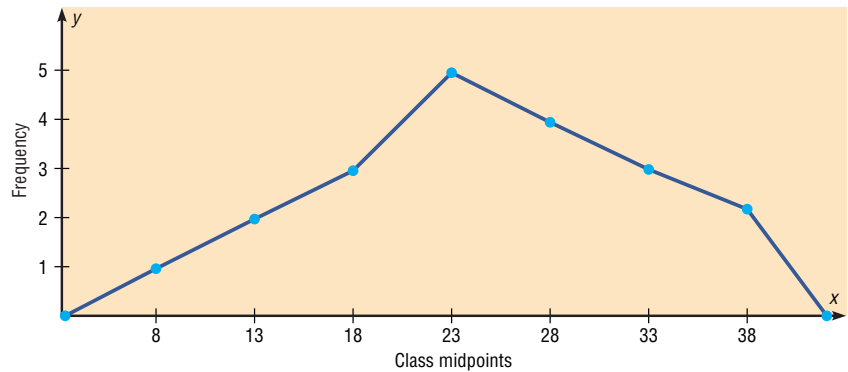
1. The histogram.
2. The frequency polygon.
3. The cumulative frequency graph, or ogive (pronounced o-jive).

An example of each type of graph is shown in Figure 2-1. The data for each graph are the distribution of the miles that 20 randomly selected runners ran during a given week.

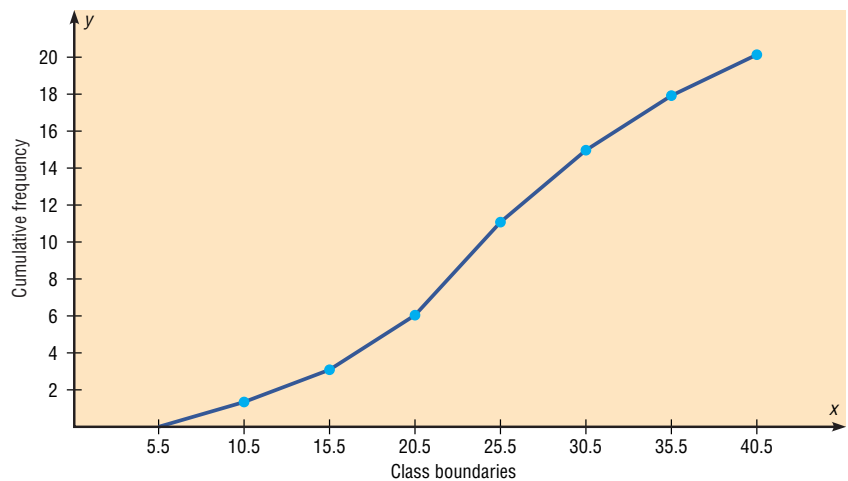
**Figure 2-1**  
Examples of Commonly  
Used Graphs



(a) Histogram



(b) Frequency polygon



(c) Cumulative frequency graph

## The Histogram

The **histogram** is a graph that displays the data by using vertical bars of various heights to represent the frequencies.

### Example 2–4

#### Historical Note

Graphs originated when ancient astronomers drew the position of the stars in the heavens. Roman surveyors also used coordinates to locate landmarks on their maps.

The development of statistical graphs can be traced to William Playfair (1748–1819), an engineer/drafter who used graphs to present economic data pictorially.

Construct a histogram to represent the data shown below for the record high temperatures for each of the 50 states (see Example 2–2).

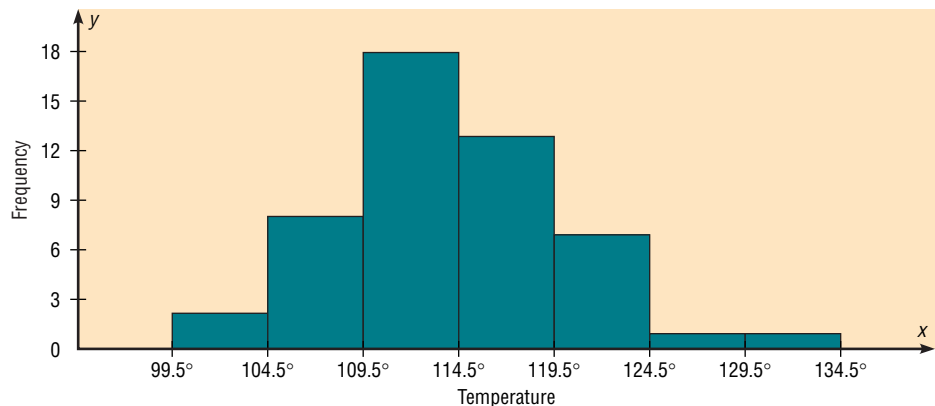
Class boundaries	Frequency
99.5–104.5	2
104.5–109.5	8
109.5–114.5	18
114.5–119.5	13
119.5–124.5	7
124.5–129.5	1
129.5–134.5	1

#### Solution

**STEP 1** Draw and label the  $x$  and  $y$  axes. The  $x$  axis is always the horizontal axis, and the  $y$  axis is always the vertical axis.

**STEP 2** Represent the frequency on the  $y$  axis and the class boundaries on the  $x$  axis.

**STEP 3** Using the frequencies as the heights, draw vertical bars for each class. See Figure 2–2.



**Figure 2–2**

Histogram for Example 2–4

As the histogram shows, the class with the greatest number of data values (18) is 109.5–114.5, followed by 13 for 114.5–119.5. The graph also has one peak with the data clustering around it.

## The Frequency Polygon

Another way to represent the same data set is by using a frequency polygon.

The **frequency polygon** is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points.



The next example shows the procedure for constructing a frequency polygon.

### Example 2–5

Using the frequency distribution given in Example 2–4, construct a frequency polygon.

#### Solution

**STEP 1** Find the midpoints of each class. Recall that midpoints are found by adding the upper and lower boundaries and dividing by 2.

$$\frac{99.5 + 104.5}{2} = 102 \qquad \frac{104.5 + 109.5}{2} = 107$$

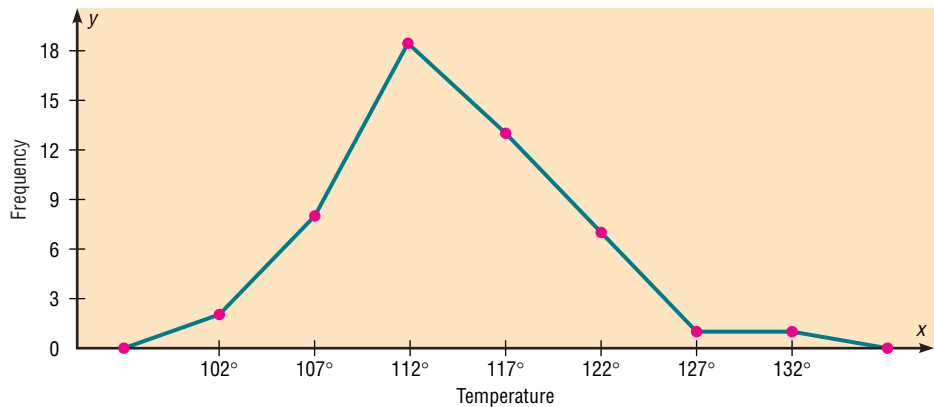
And so on. The midpoints are listed next.

Class boundaries	Midpoints	Frequency
99.5–104.5	102	2
104.5–109.5	107	8
109.5–114.5	112	18
114.5–119.5	117	13
119.5–124.5	122	7
124.5–129.5	127	1
129.5–134.5	132	1

**STEP 2** Draw the  $x$  and  $y$  axes. Label the  $x$  axis with the midpoint of each class, and then use a suitable scale on the  $y$  axis for the frequencies.

**STEP 3** Using the midpoints for the  $x$  values and the frequencies as the  $y$  values, plot the points.

**STEP 4** Connect adjacent points with straight lines. Draw a line back to the  $x$  axis at the beginning and end of the graph, at the same distance that the previous and next midpoints would be located, as shown in Figure 2–3.

**Figure 2-3**Frequency Polygon for  
Example 2-5

The frequency polygon and the histogram are two different ways to represent the same data set. The choice of which one to use is left to the discretion of the researcher.

### The Ogive

The third type of graph that can be used represents the cumulative frequencies for the classes. This type of graph is called the cumulative frequency graph or ogive. The **cumulative frequency** is the sum of the frequencies accumulated up to the upper boundary of a class in the distribution.

---

The **ogive** is a graph that represents the cumulative frequencies for the classes in a frequency distribution.

---

Example 2-6 shows the procedure for constructing an ogive.

### Example 2-6

Construct an ogive for the frequency distribution described in Example 2-4.

#### Solution

**STEP 1** Find the cumulative frequency for each class.

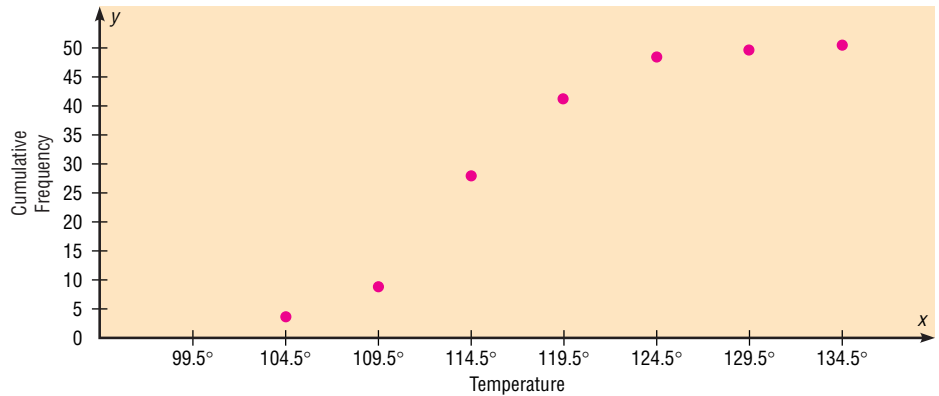
Class boundaries	Cumulative frequency
99.5–104.5	2
104.5–109.5	10
109.5–114.5	28
114.5–119.5	41
119.5–124.5	48
124.5–129.5	49
129.5–134.5	50

**STEP 2** Draw the  $x$  and  $y$  axes. Label the  $x$  axis with the class boundaries. Use an appropriate scale for the  $y$  axis to represent the cumulative frequencies. (Depending on the numbers in the cumulative frequency columns, scales such as 0, 1, 2, 3, . . . , or 5, 10, 15, 20, . . . , or 1000, 2000,

3000, . . . can be used. Do *not* label the y axis with the numbers in the cumulative frequency column.) In this example, a scale of 0, 5, 10, 15, . . . will be used.

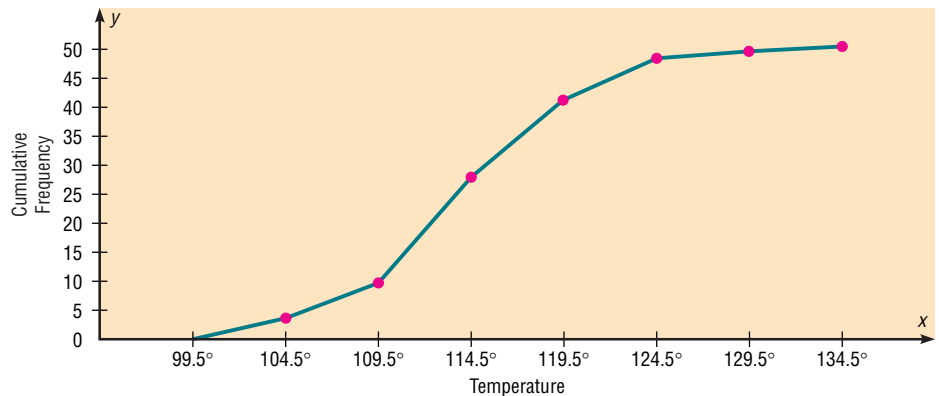
**STEP 3** Plot the cumulative frequency at each upper class boundary, as shown in Figure 2-4. Upper boundaries are used since the cumulative frequencies represent the number of data values accumulated up to the upper boundary of each class.

**Figure 2-4**  
Plotting the Cumulative Frequency for Example 2-6



**STEP 4** Starting with the first upper class boundary, 104.5, connect adjacent points with straight lines, as shown in Figure 2-5. Then extend the graph to the first lower class boundary, 99.5, on the x axis.

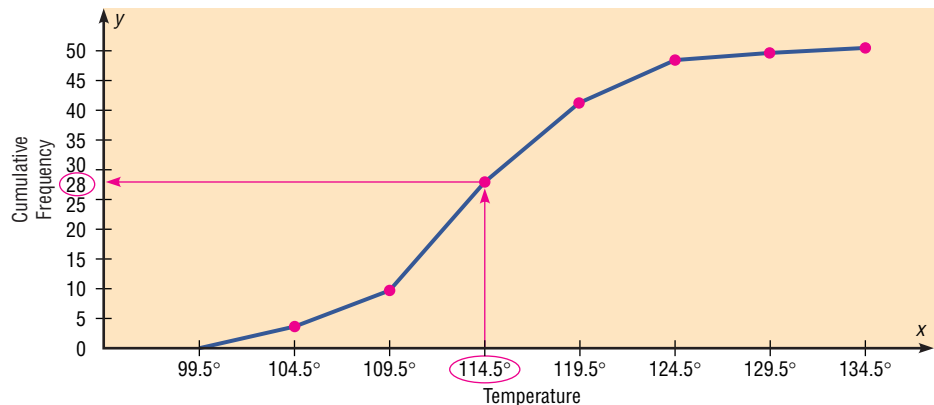
**Figure 2-5**  
Ogive for Example 2-6



Cumulative frequency graphs are used to visually represent how many values are below a certain upper class boundary. For example, to find out how many record high temperatures are less than 114.5°, locate 114.5° on the x axis, draw a vertical line up until it intersects the graph, and then draw a horizontal line at that point to the y axis. The y axis value is 28, as shown in Figure 2-6.

**Figure 2-6**

Finding a Specific  
Cumulative Frequency



The steps for drawing the three types of graphs are shown in the Procedure Table.

### Procedure Table

#### Constructing Statistical Graphs

- STEP 1** Draw and label the  $x$  and  $y$  axes.
- STEP 2** Choose a suitable scale for the frequencies or cumulative frequencies and label it on the  $y$  axis.
- STEP 3** Represent the class boundaries for the histogram or ogive, or the midpoint for the frequency polygon, on the  $x$  axis.
- STEP 4** Plot the points and then draw the bars or lines.

## Relative Frequency Graphs

The histogram, the frequency polygon, and the ogive shown previously were constructed by using frequencies in terms of the raw data. These distributions can be converted into distributions using *proportions* instead of raw data as frequencies. These types of graphs are called **relative frequency graphs**.

Graphs using relative frequencies instead of frequencies are used when the proportion of data values that fall into a given class is more important than the actual number of data values that fall into that class. For example, if one wanted to compare the age distribution of adults in the city of Philadelphia, Pennsylvania, with the age distribution of adults of Erie, Pennsylvania, one would use relative frequency distributions. The reason is that since the population of Philadelphia is 1,478,002 and the population of Erie is 105,270, the bars using the actual data values for Philadelphia would be much taller than those for the same classes for Erie.

To convert a frequency into a proportion or relative frequency, divide the frequency for each class by the total of the frequencies. The sum of the relative frequencies will always be 1. These graphs are similar to the ones that use raw data as frequencies, but the values on the  $y$  axis are in terms of proportions. The next example shows the three types of relative frequency graphs.

**Example 2–7**

Construct a histogram, frequency polygon, and ogive using relative frequencies for the distribution (shown here) of the miles 20 randomly selected runners ran during a given week.

Class boundaries	Frequency	Cumulative frequency
5.5–10.5	1	1
10.5–15.5	2	3
15.5–20.5	3	6
20.5–25.5	5	11
25.5–30.5	4	15
30.5–35.5	3	18
35.5–40.5	<u>2</u>	20
	20	

**Solution**

**STEP 1** Convert each frequency to a proportion or relative frequency by dividing the frequency for each class by the total number of observations.

For class 5.5–10.5, the relative frequency is  $\frac{1}{20} = 0.05$ .

For class 10.5–15.5, the relative frequency is  $\frac{2}{20} = 0.10$ .

For class 15.5–20.5, the relative frequency is  $\frac{3}{20} = 0.15$ .

And so on.

**STEP 2** Using the same procedure, find the relative frequencies for the cumulative frequency column. The relative frequencies are shown here.

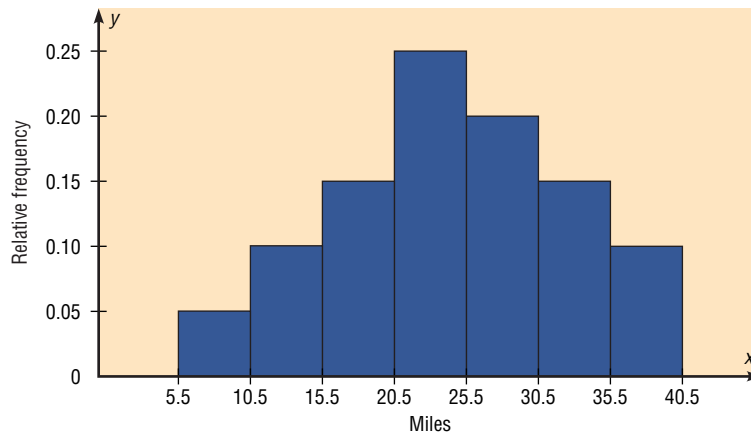
Class boundaries	Midpoints	Relative frequency	Cumulative relative frequency
5.5–10.5	8	0.05	0.05
10.5–15.5	13	0.10	0.15
15.5–20.5	18	0.15	0.30
20.5–25.5	23	0.25	0.55
25.5–30.5	28	0.20	0.75
30.5–35.5	33	0.15	0.90
35.5–40.5	38	<u>0.10</u>	1.00
		1.00	

**STEP 3** Draw each graph as shown in Figure 2–7. For the histogram and ogive, use the class boundaries along the  $x$  axis. For the frequency polygon, use the midpoints on the  $x$  axis. The scale on the  $y$  axis uses proportions.

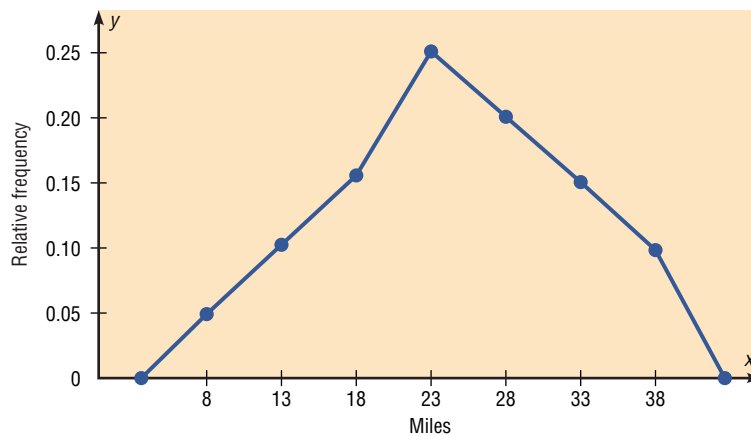
When analyzing histograms and frequency polygons, look at the shape of the curve. For example, does it have one peak or two peaks, or is it relatively flat, or is it U-shaped? Are the data values spread out on the graph, or are they clustered around the



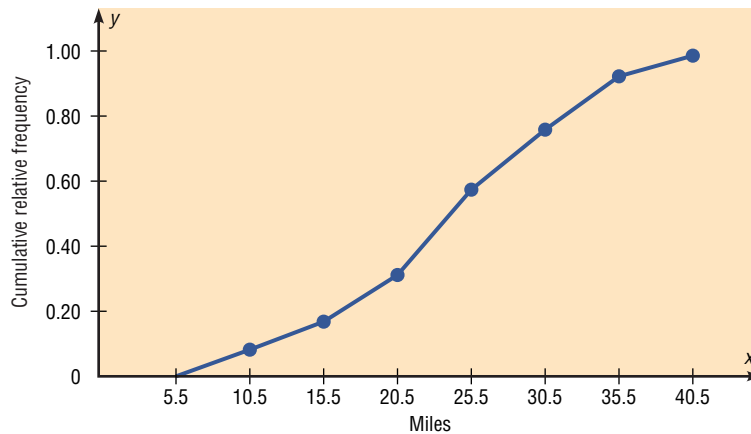
**Figure 2-7**  
 Graphs for Example 2-7



(a) Histogram



(b) Frequency polygon



(c) Ogive

center? Are there data values in the extreme ends? These may be *outliers*. (See Section 3–4 for an explanation of outliers.) Are there any gaps in the histogram, or does the frequency polygon touch the  $x$  axis somewhere other than the ends? Finally, are the data clustered at one end or the other, indicating a *skewed distribution*? (See Section 3–2 for an explanation of skewness.)

For example, the histogram for the record high temperatures shown in Figure 2–2 (page 46) shows a single peaked distribution, with the class 109.5–114.5 containing the largest number of temperatures. The distribution has no gaps, and there are fewer temperatures in the highest class than in the lowest class.

## Exercises

**2–19.** For 108 randomly selected college applicants, the following frequency distribution for entrance exam scores was obtained. Construct a histogram, frequency polygon, and ogive for the data. (The data for this exercise will be used for Exercise 2–31.)

Class limits	Frequency
90–98	6
99–107	22
108–116	43
117–125	28
126–134	9

Applicants who score above 107 need not enroll in a summer developmental program. In this group, how many students do not have to enroll in the developmental program?

**2–20.** For 75 employees of a large department store, the following distribution for years of service was obtained. Construct a histogram, frequency polygon, and ogive for the data. (The data for this exercise will be used for Exercise 2–32.)

Class limits	Frequency
1–5	21
6–10	25
11–15	15
16–20	0
21–25	8
26–30	6

A majority of the employees have worked for how many years or less?

**2–21.** Construct a histogram, frequency polygon, and ogive for the data in Exercise 2–11 and analyze the results.

**2–22.** Construct a histogram, frequency polygon, and ogive for the data in Exercise 2–12 and analyze the results.

**2–23.** Thirty automobiles were tested for fuel efficiency, in miles per gallon (mpg). The following frequency distribution was obtained. Construct a histogram, frequency polygon, and ogive for the data. (The data for this exercise will be used for Exercise 2–33.)

Class boundaries	Frequency
7.5–12.5	3
12.5–17.5	5
17.5–22.5	15
22.5–27.5	5
27.5–32.5	2

**2–24.** Construct a histogram, frequency polygon, and ogive for the data in Exercise 2–14 and analyze the results. (The data in this exercise will be used for Exercise 2–34.)

**2–25.** In a class of 35 students, the following grade distribution was found. Construct a histogram, frequency polygon, and ogive for the data. (A = 4, B = 3, C = 2, D = 1, F = 0.) (The data in this exercise will be used for Exercise 2–35.)

Grade	Frequency
0	3
1	6
2	9
3	12
4	5

A grade of C or better is required for the next level course. Were the majority of the students able to meet this requirement?

**2–26.** In a study of reaction times of dogs to a specific stimulus, an animal trainer obtained the following data, given in seconds. Construct a histogram, frequency polygon, and ogive for the data and analyze the results. (The histogram in this exercise will be used for Exercises 2–36, 3–16, and 3–68.)

Class limits	Frequency
2.3–2.9	10
3.0–3.6	12
3.7–4.3	6
4.4–5.0	8
5.1–5.7	4
5.8–6.4	2

**2–27.** Construct a histogram, frequency polygon, and ogive for the data in Exercise 2–15 and analyze the results.

**2–28.** To determine their lifetimes, 80 randomly selected batteries were tested. The following frequency distribution was obtained. The data values are in hours. Construct a histogram, frequency polygon, and ogive for the data and analyze the results.

Class boundaries	Frequency
63.5–74.5	10
74.5–85.5	15
85.5–96.5	22
96.5–107.5	17
107.5–118.5	11
118.5–129.5	5

**2–29.** Construct a histogram, frequency polygon, and ogive for the data in Exercise 2–16 and analyze the results.

**2–30.** For the data in Exercise 2–18, construct a histogram for the home run distances for each player and compare them. Are they basically the same, or are there any noticeable differences? Explain your answer.

**2–31.** For the data in Exercise 2–19, construct a histogram, frequency polygon, and ogive, using relative frequencies. What proportion of the applicants need to enroll in the summer developmental program?

**2–32.** For the data in Exercise 2–20, construct a histogram, frequency polygon, and ogive, using relative frequencies. What proportion of the employees have been with the store for more than 20 years?

**2–33.** For the data in Exercise 2–23, construct a histogram, frequency polygon, and ogive, using relative frequencies. What proportion of the automobiles had a fuel efficiency of 17.5 miles per gallon or higher?

**2–34.** For the data in Exercise 2–14, construct a histogram, frequency polygon, and ogive, using relative frequencies.

**2–35.** For the data in Exercise 2–25, construct a histogram, frequency polygon, and ogive, using relative frequencies. What proportion of the students cannot meet the requirement for enrollment in the next course?

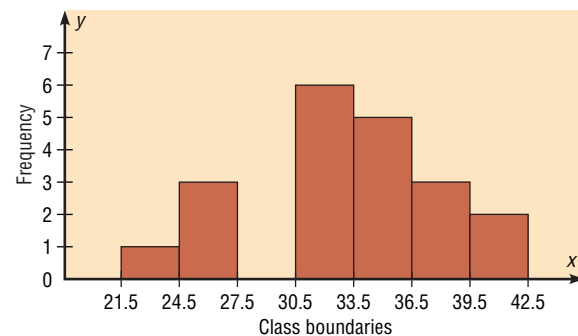
**2–36.** The animal trainer in Exercise 2–26 selected another group of dogs that were much older than the first group and measured their reaction times to the same stimulus. Construct a histogram, frequency polygon, and ogive for the data.

Class limits	Frequency
2.3–2.9	1
3.0–3.6	3
3.7–4.3	4
4.4–5.0	16
5.1–5.7	14
5.8–6.4	4

Analyze the results and compare the histogram for this group with the one obtained in Exercise 2–26. Are there any differences in the histograms? (The data in this exercise will be used for Exercise 3–16 and 3–68.)

**\*2–37.** Using the following histogram:

- Construct a frequency distribution; include class limits, class frequencies, midpoints, and cumulative frequencies.
- Construct a frequency polygon.
- Construct an ogive.



**\*2–38.** Using the results from Exercise 2–37, answer the following questions.

- How many values are in the class 27.5–30.5?
- How many values fall between 24.5 and 36.5?
- How many values are below 33.5?
- How many values are above 30.5?

## Technology Step by Step

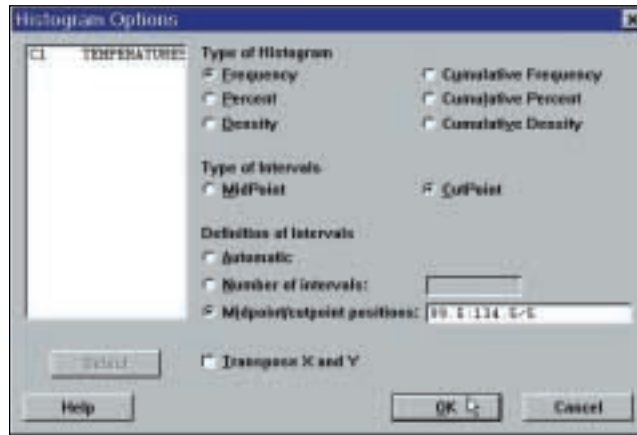
### Minitab Step by Step

### Constructing a Histogram

#### Example MT2-3

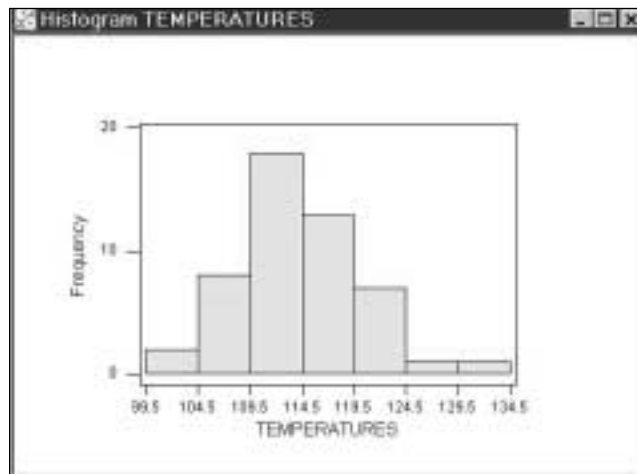
1. Enter the data from Example 2-2, the high temperatures for the 50 states.
2. Select **Graph>Histogram**.
3. In the dialog box double click C1 **TEMPERATURES**.
4. Click [Options].
  - a. In **Type of Interval**: check **CutPoint**: to use class boundaries instead of midpoints.
  - b. In **Definition of Intervals**: click the button for **Midpoint/cutpoint positions**: then type the boundaries MINITAB will use. The format is **Lowest:Highest/Increment**.

Histogram Option



5. Click [OK]. The Histogram Dialog Box will be displayed.
  6. Click [Edit Attributes]. Select **Solid** for **Fill** and **Yellow** for the **Back** color. This will make the bars of the histogram a solid yellow color.
  7. Click [OK] twice.
- The resulting graph window can be printed, copied to the clipboard or saved.

Histogram of  
Temperatures



## TI-83 Step by Step

### Constructing a Histogram

In order to display the graphs on the screen, enter the appropriate values in the calculator using the Window menu. The default values are  $X_{\min} = -10$ ,  $X_{\max} = +10$ ,  $Y_{\min} = -10$ , and  $Y_{\max} = +10$ .

The  $X_{\text{scl}}$  changes the distance between the tick marks on the  $x$  axis and can be used to change the class width for the histogram.

To change the values in the Window:

1. Press **WINDOW**.
2. Move the cursor to the value that needs to be changed. Then type in the desired value and press **ENTER**.
3. Continue until all values are appropriate.
4. Press [**2nd**] [**QUIT**] to leave the Window menu.

To plot the histogram:

1. Enter the data in  $L_1$ .
2. Make sure Window values are appropriate for the histogram.
3. Press [**2nd**] [**STAT PLOT**] **ENTER**.
4. Press **ENTER** to turn the plot on, if necessary.
5. Move cursor to the Histogram symbol and press **ENTER**, if necessary.
6. Move cursor to  $L_1$  on the Xlist. Press **ENTER**, if necessary.
7. Move cursor to 1 on Freq. Press **ENTER**, if necessary.
8. Press **GRAPH** to display the histogram.
9. To obtain various coordinates, press the **TRACE** key, followed by ◀ or ▶ keys.

#### Example TI2-1

Plot a histogram for the following data from Examples 2-2 and 2-4.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Set the Window values as follows:

$$X_{\min} = 100$$

$$X_{\max} = 135$$

$$Y_{\min} = -5$$

$$Y_{\max} = 20$$

Input

```

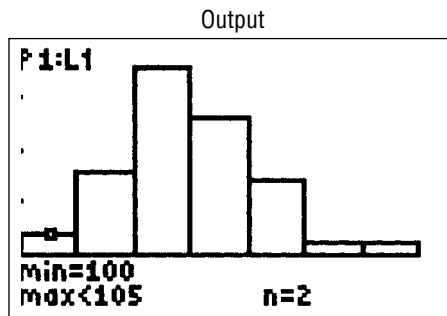
WINDOW
Xmin=100
Xmax=135
Xscl=5
Ymin=-5
Ymax=20
Yscl=5
Xres=1
  
```

Input

```

PLOT1 Plot2 Plot3
Off Off
Type: [L1] [L2] [L3]
Xlist:L1
Freq:1
  
```

Press **TRACE** and use the arrow keys to determine the number of values in each group.



## Excel Step by Step

### Constructing a Histogram

#### Example XL2-1

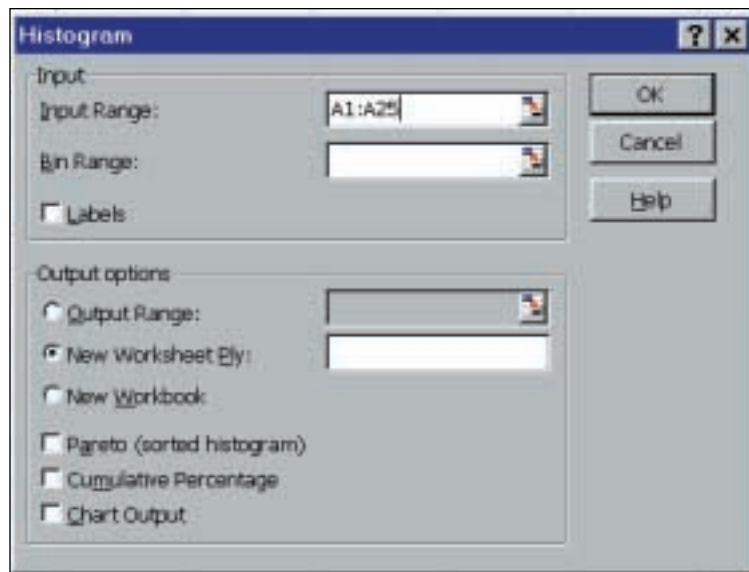
To make a histogram:

1. Press **[Ctrl]-N** for a new worksheet.
2. Enter the data below in column A, one number per cell.

68	80	69	81	72	100	101	73	102	93
91	92	93	88	82	83	75	75	89	96
103	83	84	85	89					

3. Select **Tools > Data Analysis**.
4. In **Data Analysis**, select **Histogram** and click the **[OK]** button.
5. In the **Histogram** dialog box, type **A1:A25** as the **Input Range**.

Histogram Dialog Box



6. Select **New Worksheet Ply** and **Chart Output**. Click **[OK]**.

Excel presents both a table and a chart on the new worksheet ply. It decides “bins” for the histogram itself (here it picked a bin size of seven units), but you can also define your own bin range on the data worksheet. Here is the histogram with Excel-selected bins.

Histogram for  
Example XL2-1

