Two types of frequency distributions that are most often used are the *categorical frequency distribution* and the *grouped frequency distribution.* The procedures for constructing these distributions are shown now.

Categorical Frequency Distributions

The **categorical frequency distribution** is used for data that can be placed in specific categories, such as nominal- or ordinal-level data. For example, data such as political affiliation, religious affiliation, or major field of study would use categorical frequency distributions.

| Example 2–1 | Distribution of Blood Types |

Twenty-five army inductees were given a blood test to determine their blood type. The data set is

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Construct a frequency distribution for the data.

Solution

Since the data are categorical, discrete classes can be used. There are four blood types: A, B, O, and AB. These types will be used as the classes for the distribution.

The procedure for constructing a frequency distribution for categorical data is given next.

Step 1   Make a table as shown.

A Class	B Tally	C Frequency	D Percent
A			
B			
O			
AB			

Step 2   Tally the data and place the results in column B.

Step 3   Count the tallies and place the results in column C.

Step 4   Find the percentage of values in each class by using the formula

$$\% = \frac{f}{n} \cdot 100\%$$

where f = frequency of the class and n = total number of values. For example, in the class of type A blood, the percentage is

$$\% = \frac{5}{25} \cdot 100\% = 20\%$$

Percentages are not normally part of a frequency distribution, but they can be added since they are used in certain types of graphs such as pie graphs. Also, the decimal equivalent of a percent is called a *relative frequency.*

Step 5   Find the totals for columns C (frequency) and D (percent). The completed table is shown.

A	B	C	D
Class	**Tally**	**Frequency**	**Percent**
A	𝍤	5	20
B	𝍤 //	7	28
O	𝍤 ////	9	36
AB	////	4	16
		Total 25	100

For the sample, more people have type O blood than any other type.

Grouped Frequency Distributions

When the range of the data is large, the data must be grouped into classes that are more than one unit in width, in what is called a **grouped frequency distribution.** For example, a distribution of the number of hours that boat batteries lasted is the following.

Class limits	Class boundaries	Tally	Frequency
24–30	23.5–30.5	///	3
31–37	30.5–37.5	/	1
38–44	37.5–44.5	𝍤	5
45–51	44.5–51.5	𝍤 ////	9
52–58	51.5–58.5	𝍤 /	6
59–65	58.5–65.5	/	1
			25

The procedure for constructing the preceding frequency distribution is given in Example 2–2; however, several things should be noted. In this distribution, the values 24 and 30 of the first class are called *class limits.* The **lower class limit** is 24; it represents the smallest data value that can be included in the class. The **upper class limit** is 30; it represents the largest data value that can be included in the class. The numbers in the second column are called **class boundaries.** These numbers are used to separate the classes so that there are no gaps in the frequency distribution. The gaps are due to the limits; for example, there is a gap between 30 and 31.

Students sometimes have difficulty finding class boundaries when given the class limits. The basic rule of thumb is that *the class limits should have the same decimal place value as the data, but the class boundaries should have one additional place value and end in a 5.* For example, if the values in the data set are whole numbers, such as 24, 32, and 18, the limits for a class might be 31–37, and the boundaries are 30.5–37.5. Find the boundaries by subtracting 0.5 from 31 (the lower class limit) and adding 0.5 to 37 (the upper class limit).

$$\text{Lower limit} - 0.5 = 31 - 0.5 = 30.5 = \text{lower boundary}$$

$$\text{Upper limit} + 0.5 = 37 + 0.5 = 37.5 = \text{upper boundary}$$

If the data are in tenths, such as 6.2, 7.8, and 12.6, the limits for a class hypothetically might be 7.8–8.8, and the boundaries for that class would be 7.75–8.85. Find these values by subtracting 0.05 from 7.8 and adding 0.05 to 8.8.

Finally, the **class width** for a class in a frequency distribution is found by subtracting the lower (or upper) class limit of one class from the lower (or upper) class limit of the next class. For example, the class width in the preceding distribution on the duration of boat batteries is 7, found from $31 - 24 = 7$.

The class width can also be found by subtracting the lower boundary from the upper boundary for any given class. In this case, $30.5 - 23.5 = 7$.

Note: Do not subtract the limits of a single class. It will result in an incorrect answer.

The researcher must decide how many classes to use and the width of each class. To construct a frequency distribution, follow these rules:

1. *There should be between 5 and 20 classes.* Although there is no hard-and-fast rule for the number of classes contained in a frequency distribution, it is of the utmost importance to have enough classes to present a clear description of the collected data.

2. *It is preferable but not absolutely necessary that the class width be an odd number.* This ensures that the midpoint of each class has the same place value as the data. The **class midpoint** X_m is obtained by adding the lower and upper boundaries and dividing by 2, or adding the lower and upper limits and dividing by 2:

$$X_m = \frac{\text{lower boundary} + \text{upper boundary}}{2}$$

or

$$X_m = \frac{\text{lower limit} + \text{upper limit}}{2}$$

For example, the midpoint of the first class in the example with boat batteries is

$$\frac{24 + 30}{2} = 27 \qquad \text{or} \qquad \frac{23.5 + 30.5}{2} = 27$$

The midpoint is the numeric location of the center of the class. Midpoints are necessary for graphing (see Section 2–2). If the class width is an even number, the midpoint is in tenths. For example, if the class width is 6 and the boundaries are 5.5 and 11.5, the midpoint is

$$\frac{5.5 + 11.5}{2} = \frac{17}{2} = 8.5$$

Rule 2 is only a suggestion, and it is not rigorously followed, especially when a computer is used to group data.

3. *The classes must be mutually exclusive.* Mutually exclusive classes have nonoverlapping class limits so that data cannot be placed into two classes. Many times, frequency distributions such as

Age
10–20
20–30
30–40
40–50

are found in the literature or in surveys. If a person is 40 years old, into which class should she or he be placed? A better way to construct a frequency distribution is to use classes such as

Age
10–20
21–31
32–42
43–53

4. *The classes must be continuous.* Even if there are no values in a class, the class must be included in the frequency distribution. There should be no gaps in a

frequency distribution. The only exception occurs when the class with a zero frequency is the first or last class. A class with a zero frequency at either end can be omitted without affecting the distribution.

5. *The classes must be exhaustive.* There should be enough classes to accommodate all the data.

6. *The classes must be equal in width.* This avoids a distorted view of the data.

   One exception occurs when a distribution has a class that is open-ended. That is, the class has no specific beginning value or no specific ending value. A frequency distribution with an open-ended class is called an **open-ended distribution.** Here are two examples of distributions with open-ended classes.

Age	Frequency		Minutes	Frequency
10–20	3		Below 110	16
21–31	6		110–114	24
32–42	4		115–119	38
43–53	10		120–124	14
54 and above	8		125–129	5

The frequency distribution for age is open-ended for the last class, which means that anybody who is 54 years or older will be tallied in the last class. The distribution for minutes is open-ended for the first class, meaning that any minute values below 110 will be tallied in that class.

Example 2–2 shows the procedure for constructing a grouped frequency distribution, i.e., when the classes contain more than one data value.

Example 2–2

Record High Temperatures

These data represent the record high temperatures in degrees Fahrenheit (°F) for each of the 50 states. Construct a grouped frequency distribution for the data using 7 classes.

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

Source: *The World Almanac and Book of Facts.*

Solution

The procedure for constructing a grouped frequency distribution for numerical data follows.

Step 1   Determine the classes.

Find the highest value and lowest value: $H = 134$ and $L = 100$.

Find the range: $R =$ highest value $-$ lowest value $= H - L$, so

$R = 134 - 100 = 34$

Select the number of classes desired (usually between 5 and 20). In this case, 7 is arbitrarily chosen.

Find the class width by dividing the range by the number of classes.

$$\text{Width} = \frac{R}{\text{number of classes}} = \frac{34}{7} = 4.9$$

Round the answer up to the nearest whole number if there is a remainder: $4.9 \approx 5$. (Rounding *up* is different from rounding *off*. A number is rounded up if there is any decimal remainder when dividing. For example, $85 \div 6 = 14.167$ and is rounded up to 15. Also, $53 \div 4 = 13.25$ and is rounded up to 14. Also, after dividing, if there is no remainder, you will need to add an extra class to accommodate all the data.)

Select a starting point for the lowest class limit. This can be the smallest data value or any convenient number less than the smallest data value. In this case, 100 is used. Add the width to the lowest score taken as the starting point to get the lower limit of the next class. Keep adding until there are 7 classes, as shown, 100, 105, 110, etc.

Subtract one unit from the lower limit of the second class to get the upper limit of the first class. Then add the width to each upper limit to get all the upper limits.

$$105 - 1 = 104$$

The first class is 100–104, the second class is 105–109, etc.

Find the class boundaries by subtracting 0.5 from each lower class limit and adding 0.5 to each upper class limit:

99.5–104.5, 104.5–109.5, etc.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies.

The completed frequency distribution is

Class limits	Class boundaries	Tally	Frequency
100–104	99.5–104.5	//	2
105–109	104.5–109.5	⟍⟍ ///	8
110–114	109.5–114.5	⟍⟍ ⟍⟍ ⟍⟍ ///	18
115–119	114.5–119.5	⟍⟍ ⟍⟍ ///	13
120–124	119.5–124.5	⟍⟍ //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			$n = \Sigma f = 50$

The frequency distribution shows that the class 109.5–114.5 contains the largest number of temperatures (18) followed by the class 114.5–119.5 with 13 temperatures. Hence, most of the temperatures (31) fall between 109.5 and 119.5°F.

Sometimes it is necessary to use a *cumulative frequency distribution*. A **cumulative frequency distribution** is a distribution that shows the number of data values less than or equal to a specific value (usually an upper boundary). The values are found by adding the frequencies of the classes less than or equal to the upper class boundary of a specific class. This gives an ascending cumulative frequency. In this example, the cumulative frequency for the first class is $0 + 2 = 2$; for the second class it is $0 + 2 + 8 = 10$; for the third class it is $0 + 2 + 8 + 18 = 28$. Naturally, a shorter way to do this would be to just add the cumulative frequency of the class below to the frequency of the given class. For

example, the cumulative frequency for the number of data values less than 114.5 can be found by adding $10 + 18 = 28$. The cumulative frequency distribution for the data in this example is as follows:

	Cumulative frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Cumulative frequencies are used to show how many data values are accumulated up to and including a specific class. In Example 2–2, 28 of the total record high temperatures are less than or equal to 114°F. Forty-eight of the total record high temperatures are less than or equal to 124°F.

After the raw data have been organized into a frequency distribution, it will be analyzed by looking for peaks and extreme values. The peaks show which class or classes have the most data values compared to the other classes. Extreme values, called outliers, show large or small data values that are relative to other data values.

When the range of the data values is relatively small, a frequency distribution can be constructed using single data values for each class. This type of distribution is called an **ungrouped frequency distribution** and is shown next.

Example 2–3

MPGs for SUVs

The data shown here represent the number of miles per gallon (mpg) that 30 selected four-wheel-drive sports utility vehicles obtained in city driving. Construct a frequency distribution, and analyze the distribution.

12	17	12	14	16	18
16	18	12	16	17	15
15	16	12	15	16	16
12	14	15	12	15	15
19	13	16	18	16	14

Source: *Model Year Fuel Economy Guide.* United States Environmental Protection Agency.

Solution

Step 1 Determine the classes. Since the range of the data set is small ($19 - 12 = 7$), classes consisting of a single data value can be used. They are 12, 13, 14, 15, 16, 17, 18, 19.

Note: If the data are continuous, class boundaries can be used. Subtract 0.5 from each class value to get the lower class boundary, and add 0.5 to each class value to get the upper class boundary.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies, and find the cumulative frequencies.

The completed ungrouped frequency distribution is

Class limits	Class boundaries	Tally	Frequency
12	11.5–12.5	ⷮⷛ /	6
13	12.5–13.5	/	1
14	13.5–14.5	///	3
15	14.5–15.5	ⷮⷛ /	6
16	15.5–16.5	ⷮⷛ ///	8
17	16.5–17.5	//	2
18	17.5–18.5	///	3
19	18.5–19.5	/	1

In this case, almost one-half (14) of the vehicles get 15 or 16 miles per gallon. The cumulative frequencies are

	Cumulative frequency
Less than 11.5	0
Less than 12.5	6
Less than 13.5	7
Less than 14.5	10
Less than 15.5	16
Less than 16.5	24
Less than 17.5	26
Less than 18.5	29
Less than 19.5	30

The steps for constructing a grouped frequency distribution are summarized in the following Procedure Table.

Procedure Table

Constructing a Grouped Frequency Distribution

Step 1   Determine the classes.

Find the highest and lowest values.

Find the range.

Select the number of classes desired.

Find the width by dividing the range by the number of classes and rounding up.

Select a starting point (usually the lowest value or any convenient number less than the lowest value); add the width to get the lower limits.

Find the upper class limits.

Find the boundaries.

Step 2   Tally the data.

Step 3   Find the numerical frequencies from the tallies, and find the cumulative frequencies.

When you are constructing a frequency distribution, the guidelines presented in this section should be followed. However, you can construct several different but correct frequency distributions for the same data by using a different class width, a different number of classes, or a different starting point.

Furthermore, the method shown here for constructing a frequency distribution is not unique, and there are other ways of constructing one. Slight variations exist, especially in computer packages. But regardless of what methods are used, classes should be mutually exclusive, continuous, exhaustive, and of equal width.

In summary, the different types of frequency distributions were shown in this section. The first type, shown in Example 2–1, is used when the data are categorical (nominal), such as blood type or political affiliation. This type is called a categorical frequency distribution. The second type of distribution is used when the range is large and classes several units in width are needed. This type is called a grouped frequency distribution and is shown in Example 2–2. Another type of distribution is used for numerical data and when the range of data is small, as shown in Example 2–3. Since each class is only one unit, this distribution is called an ungrouped frequency distribution.

All the different types of distributions are used in statistics and are helpful when one is organizing and presenting data.

The reasons for constructing a frequency distribution are as follows:

1. To organize the data in a meaningful, intelligible way.
2. To enable the reader to determine the nature or shape of the distribution.
3. To facilitate computational procedures for measures of average and spread (shown in Sections 3–1 and 3–2).
4. To enable the researcher to draw charts and graphs for the presentation of data (shown in Section 2–2).
5. To enable the reader to make comparisons among different data sets.

The factors used to analyze a frequency distribution are essentially the same as those used to analyze histograms and frequency polygons, which are shown in Section 2–2.

Applying the Concepts **2–1**

Ages of Presidents at Inauguration

The data represent the ages of our Presidents at the time they were first inaugurated.

57	61	57	57	58	57	61	54	68
51	49	64	50	48	65	52	56	46
54	49	51	47	55	55	54	42	51
56	55	51	54	51	60	62	43	55
56	61	52	69	64	46	54	47	

1. Were the data obtained from a population or a sample? Explain your answer.
2. What was the age of the oldest President?
3. What was the age of the youngest President?
4. Construct a frequency distribution for the data. (Use your own judgment as to the number of classes and class size.)
5. Are there any peaks in the distribution?

6. ldentify any possible outliers.

7. Write a brief summary of the nature of the data as shown in the frequency distribution.

See page 101 for the answers.

Answers not appearing on the page can be found in the answers appendix.

Exercises 2–1

1. List five reasons for organizing data into a frequency distribution.

2. Name the three types of frequency distributions, and explain when each should be used. Categorical, ungrouped, grouped

3. Find the class boundaries, midpoints, and widths for each class.

a. 32–38 31.5–38.5, 35, 7

b. 86–104 85.5–104.5, 95, 19

c. 895–905 894.5–905.5, 900, 11

d. 12.3–13.5 12.25–13.55, 12.9, 1.3

e. 3.18–4.96 3.175–4.965, 4.07, 1.79

4. How many classes should frequency distributions have? Why should the class width be an odd number?

5. Shown here are four frequency distributions. Each is incorrectly constructed. State the reason why.

a.
Class	Frequency
27–32	1
33–38	0
39–44	6
45–49	4
50–55	2

Class width is not uniform.

b.
Class	Frequency
5–9	1
9–13	2
13–17	5
17–20	6
20–24	3

Class limits overlap, and class width is not uniform.

c.
Class	Frequency
123–127	3
128–132	7
138–142	2
143–147	19

A class has been omitted.

d.
Class	Frequency
9–13	1
14–19	6
20–25	2
26–28	5
29–32	9

Class width is not uniform.

6. What are open-ended frequency distributions? Why are they necessary?

7. Trust in Internet Information A survey was taken on how much trust people place in the information they read on the Internet. Construct a categorical frequency distribution for the data. A = trust in everything they read, M = trust in most of what they read, H = trust in about one-half of what they read, S = trust in a small portion of what they read. (Based on information from the *UCLA Internet Report.*)

```
M  M  M  A  H  M  S  M  H  M
S  M  M  M  M  A  M  M  A  M
M  M  H  M  M  M  H  M  H  M
A  M  M  M  H  M  M  M  M  M
```

8. Grams per Food Serving The data shown are the number of grams per serving of 30 selected brands of cakes. Construct a frequency distribution using 5 classes.

```
32  47  51  41  46  30
46  38  34  34  52  48
48  38  43  41  21  24
25  29  33  45  51  32
32  27  23  23  34  35
```

Source: *The Complete Food Counts.*

9. Weights of the NBA's Top 50 Players Listed are the weights of the NBA's top 50 players. Construct a grouped frequency distribution and a cumulative frequency distribution with 8 classes. Analyze the results in terms of peaks, extreme values, etc.

```
240  210  220  260  250  195  230  270  325  225
165  295  205  230  250  210  220  210  230  202
250  265  230  210  240  245  225  180  175  215
215  235  245  250  215  210  195  240  240  225
260  210  190  260  230  190  210  230  185  260
```

Source: www.msn.foxsports.com

10. Stories in the World's Tallest Buildings The number of stories in each of the world's 30 tallest buildings follows. Construct a grouped frequency distribution and a cumulative frequency distribution with 7 classes.