Statistics Today

How Long Are You Delayed by Road Congestion?

No matter where you live, at one time or another, you have been stuck in traffic. To see whether there are more traffic delays in some cities than in others, statisticians make comparisons using descriptive statistics. A statistical study by the Texas Transportation Institute found that a driver is delayed by road congestion an average of 36 hours per year. To see how selected cities compare to this average, see Statistics Today—Revisited at the end of the chapter.

This chapter will show you how to obtain and interpret descriptive statistics such as measures of average, measures of variation, and measures of position.

Introduction

Chapter 2 showed how you can gain useful information from raw data by organizing them into a frequency distribution and then presenting the data by using various graphs. This chapter shows the statistical methods that can be used to summarize data. The most familiar of these methods is the finding of averages.

For example, you may read that the average speed of a car crossing midtown Manhattan during the day is 5.3 miles per hour or that the average number of minutes an American father of a 4-year-old spends alone with his child each day is 42.[1]

In the book *American Averages* by Mike Feinsilber and William B. Meed, the authors state:

> *"Average" when you stop to think of it is a funny concept. Although it describes all of us it describes none of us. . . . While none of us wants to be the average American, we all want to know about him or her.*

Interesting Fact

A person has on average 1460 dreams in 1 year.

The authors go on to give examples of averages:

> *The average American man is five feet, nine inches tall; the average woman is five feet, 3.6 inches.*
> *The average American is sick in bed seven days a year missing five days of work.*
> *On the average day, 24 million people receive animal bites.*
> *By his or her 70th birthday, the average American will have eaten 14 steers, 1050 chickens, 3.5 lambs, and 25.2 hogs.*[2]

[1]"Harper's Index," *Harper's* magazine.

[2]Mike Feinsilber and William B. Meed, *American Averages* (New York: Bantam Doubleday Dell).

In these examples, the word *average* is ambiguous, since several different methods can be used to obtain an average. Loosely stated, the average means the center of the distribution or the most typical case. Measures of average are also called *measures of central tendency* and include the *mean, median, mode,* and *midrange.*

Knowing the average of a data set is not enough to describe the data set entirely. Even though a shoe store owner knows that the average size of a man's shoe is size 10, she would not be in business very long if she ordered only size 10 shoes.

As this example shows, in addition to knowing the average, you must know how the data values are dispersed. That is, do the data values cluster around the mean, or are they spread more evenly throughout the distribution? The measures that determine the spread of the data values are called *measures of variation,* or *measures of dispersion.* These measures include the *range, variance,* and *standard deviation.*

Finally, another set of measures is necessary to describe data. These measures are called *measures of position.* They tell where a specific data value falls within the data set or its relative position in comparison with other data values. The most common position measures are *percentiles, deciles,* and *quartiles.* These measures are used extensively in psychology and education. Sometimes they are referred to as *norms.*

The measures of central tendency, variation, and position explained in this chapter are part of what is called *traditional statistics.*

Section 3–4 shows the techniques of what is called *exploratory data analysis.* These techniques include the *boxplot* and the *five-number summary.* They can be used to explore data to see what they show (as opposed to the traditional techniques, which are used to confirm conjectures about the data).

3–1 Measures of Central Tendency

Chapter 1 stated that statisticians use samples taken from populations; however, when populations are small, it is not necessary to use samples since the entire population can be used to gain information. For example, suppose an insurance manager wanted to know the average weekly sales of all the company's representatives. If the company employed a large number of salespeople, say, nationwide, he would have to use a sample and make

Objective 1

Summarize data, using measures of central tendency, such as the mean, median, mode, and midrange.

an inference to the entire sales force. But if the company had only a few salespeople, say, only 87 agents, he would be able to use all representatives' sales for a randomly chosen week and thus use the entire population.

Measures found by using all the data values in the population are called *parameters.* Measures obtained by using the data values from samples are called *statistics;* hence, the average of the sales from a sample of representatives is a *statistic,* and the average of sales obtained from the entire population is a *parameter.*

A **statistic** is a characteristic or measure obtained by using the data values from a sample.

A **parameter** is a characteristic or measure obtained by using all the data values from a specific population.

Historical Note

In 1796, Adolphe Quetelet investigated the characteristics (heights, weights, etc.) of French conscripts to determine the "average man." Florence Nightingale was so influenced by Quetelet's work that she began collecting and analyzing medical records in the military hospitals during the Crimean War. Based on her work, hospitals began keeping accurate records on their patients.

These concepts as well as the symbols used to represent them will be explained in detail in this chapter.

General Rounding Rule In statistics the basic rounding rule is that when computations are done in the calculation, rounding should not be done until the final answer is calculated. When rounding is done in the intermediate steps, it tends to increase the difference between that answer and the exact one. But in the textbook and solutions manual, it is not practical to show long decimals in the intermediate calculations; hence, the values in the examples are carried out to enough places (usually three or four) to obtain the same answer that a calculator would give after rounding on the last step.

The Mean

The *mean,* also known as the *arithmetic average,* is found by adding the values of the data and dividing by the total number of values. For example, the mean of 3, 2, 6, 5, and 4 is found by adding $3 + 2 + 6 + 5 + 4 = 20$ and dividing by 5; hence, the mean of the data is $20 \div 5 = 4$. The values of the data are represented by X's. In this data set, $X_1 = 3$, $X_2 = 2$, $X_3 = 6$, $X_4 = 5$, and $X_5 = 4$. To show a sum of the total X values, the symbol Σ (the capital Greek letter sigma) is used, and ΣX means to find the sum of the X values in the data set. The summation notation is explained in Appendix A.

The **mean** is the sum of the values, divided by the total number of values. The symbol $\overline{X}$ represents the sample mean.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\Sigma X}{n}$$

where n represents the total number of values in the sample.

For a population, the Greek letter μ (mu) is used for the mean.

$$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\Sigma X}{N}$$

where N represents the total number of values in the population.

In statistics, Greek letters are used to denote parameters, and Roman letters are used to denote statistics. Assume that the data are obtained from samples unless otherwise specified.

Example 3–1

Days Off per Year

The data represent the number of days off per year for a sample of individuals selected from nine different countries. Find the mean.

20, 26, 40, 36, 23, 42, 35, 24, 30

Source: World Tourism Organization.

Solution

$$\bar{X} = \frac{\Sigma X}{n} = \frac{20 + 26 + 40 + 36 + 23 + 42 + 35 + 24 + 30}{9} = \frac{276}{9} = 30.7 \text{ days}$$

Hence, the mean of the number of days off is 30.7 days.

Example 3–2

Hospital Infections

The data show the number of patients in a sample of six hospitals who acquired an infection while hospitalized. Find the mean.

110 76 29 38 105 31

Source: Pennsylvania Health Care Cost Containment Council.

Solution

$$\bar{X} = \frac{\Sigma X}{n} = \frac{110 + 76 + 29 + 38 + 105 + 31}{6} = \frac{389}{6} = 64.8$$

The mean of the number of hospital infections for the six hospitals is 64.8.

The mean, in most cases, is not an actual data value.

Rounding Rule for the Mean The mean should be rounded to one more decimal place than occurs in the raw data. For example, if the raw data are given in whole numbers, the mean should be rounded to the nearest tenth. If the data are given in tenths, the mean should be rounded to the nearest hundredth, and so on.

The procedure for finding the mean for grouped data uses the midpoints of the classes. This procedure is shown next.

Example 3–3

Miles Run per Week

Using the frequency distribution for Example 2–7, find the mean. The data represent the number of miles run during one week for a sample of 20 runners.

Solution

The procedure for finding the mean for grouped data is given here.

Step 1 Make a table as shown.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$
5.5–10.5	1		
10.5–15.5	2		
15.5–20.5	3		
20.5–25.5	5		
25.5–30.5	4		
30.5–35.5	3		
35.5–40.5	2		
	$n = 20$		

Interesting Fact

The average time it takes a person to find a new job is 5.9 months.

Step 2 Find the midpoints of each class and enter them in column C.

$$X_m = \frac{5.5 + 10.5}{2} = 8 \qquad \frac{10.5 + 15.5}{2} = 13 \qquad \text{etc.}$$

Step 3 For each class, multiply the frequency by the midpoint, as shown, and place the product in column D.

$1 \cdot 8 = 8 \qquad 2 \cdot 13 = 26 \qquad$ etc.

The completed table is shown here.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$
5.5–10.5	1	8	8
10.5–15.5	2	13	26
15.5–20.5	3	18	54
20.5–25.5	5	23	115
25.5–30.5	4	28	112
30.5–35.5	3	33	99
35.5–40.5	2	38	76
	$n = 20$		$\Sigma f \cdot X_m = 490$

Unusual Stat

A person looks, on average, at about 14 homes before he or she buys one.

Step 4 Find the sum of column D.

Step 5 Divide the sum by n to get the mean.

$$\overline{X} = \frac{\Sigma f \cdot X_m}{n} = \frac{490}{20} = 24.5 \text{ miles}$$

The procedure for finding the mean for grouped data assumes that the mean of all the raw data values in each class is equal to the midpoint of the class. In reality, this is not true, since the average of the raw data values in each class usually will not be exactly equal to the midpoint. However, using this procedure will give an acceptable approximation of the mean, since some values fall above the midpoint and other values fall below the midpoint for each class, and the midpoint represents an estimate of all values in the class.

The steps for finding the mean for grouped data are summarized in the next Procedure Table.

Procedure Table

Finding the Mean for Grouped Data

Step 1 Make a table as shown.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$

Step 2 Find the midpoints of each class and place them in column C.

Step 3 Multiply the frequency by the midpoint for each class, and place the product in column D.

Step 4 Find the sum of column D.

Step 5 Divide the sum obtained in column D by the sum of the frequencies obtained in column B.

The formula for the mean is

$$\overline{X} = \frac{\Sigma f \cdot X_m}{n}$$

[*Note:* The symbols $\Sigma f \cdot X_m$ mean to find the sum of the product of the frequency (f) and the midpoint (X_m) for each class.]

Speaking of Statistics

Ages of the Top 50 Wealthiest People

The histogram shows the ages of the top 50 wealthiest individuals according to *Forbes Magazine* for a recent year. The mean age is 66.04 years. The median age is 68 years. Explain why these two statistics are not enough to adequately describe the data.

The Median

Historical Note

The concept of median was used by Gauss at the beginning of the 19th century and introduced as a statistical concept by Francis Galton around 1874. The mode was first used by Karl Pearson in 1894.

An article recently reported that the median income for college professors was $43,250. This measure of central tendency means that one-half of all the professors surveyed earned more than $43,250, and one-half earned less than $43,250.

The *median* is the halfway point in a data set. Before you can find this point, the data must be arranged in order. When the data set is ordered, it is called a **data array.** The median either will be a specific value in the data set or will fall between two values, as shown in Examples 3–4 through 3–8.

The **median** is the midpoint of the data array. The symbol for the median is MD.

Steps in computing the median of a data array

Step 1 Arrange the data in order.

Step 2 Select the middle point.

Example 3–4

Hotel Rooms

The number of rooms in the seven hotels in downtown Pittsburgh is 713, 300, 618, 595, 311, 401, and 292. Find the median.

Source: Interstate Hotels Corporation.

Solution

Step 1 Arrange the data in order.

292, 300, 311, 401, 595, 618, 713

Step 2 Select the middle value.

292, 300, 311, 401, 595, 618, 713

↑

Median

Hence, the median is 401 rooms.

Example 3–5

National Park Vehicle Pass Costs

Find the median for the daily vehicle pass charge for five U.S. National Parks. The costs are $25, $15, $15, $20, and $15.

Source: National Park Service.

Solution

$15 $15 $15 $20 $25

↑

Median

The median cost is $15.

Examples 3–4 and 3–5 each had an odd number of values in the data set; hence, the median was an actual data value. When there are an even number of values in the data set, the median will fall between two given values, as illustrated in Examples 3–6, 3–7, and 3–8.

Example 3–6

Tornadoes in the United States

The number of tornadoes that have occurred in the United States over an 8-year period follows. Find the median.

684, 764, 656, 702, 856, 1133, 1132, 1303

Source: *The Universal Almanac.*

Solution

656, 684, 702, 764, 856, 1132, 1133, 1303

↑

Median

Since the middle point falls halfway between 764 and 856, find the median MD by adding the two values and dividing by 2.

$$\text{MD} = \frac{764 + 856}{2} = \frac{1620}{2} = 810$$

The median number of tornadoes is 810.

Example 3–7

Asthma Cases

The number of children with asthma during a specific year in seven local districts is shown. Find the median.

253, 125, 328, 417, 201, 70, 90

Source: Pennsylvania Department of Health.

Solution

70, 90, 125, 201, 253, 328, 417
↑
Median

Since the number 201 is at the center of the distribution, the median is 201.

Example 3–8

Magazines Purchased

Six customers purchased these numbers of magazines: 1, 7, 3, 2, 3, 4. Find the median.

Solution

1, 2, 3, 3, 4, 7 $\qquad$ $\text{MD} = \dfrac{3 + 3}{2} = 3$
↑
Median

Hence, the median number of magazines purchased is 3.

The Mode

The third measure of average is called the *mode.* The mode is the value that occurs most often in the data set. It is sometimes said to be the most typical case.

The value that occurs most often in a data set is called the **mode.**

A data set that has only one value that occurs with the greatest frequency is said to be **unimodal.**

If a data set has two values that occur with the same greatest frequency, both values are considered to be the mode and the data set is said to be **bimodal.** If a data set has more than two values that occur with the same greatest frequency, each value is used as the mode, and the data set is said to be **multimodal.** When no data value occurs more than once, the data set is said to have *no mode.* A data set can have more than one mode or no mode at all. These situations will be shown in some of the examples that follow.

Example 3–9

NFL Signing Bonuses

Find the mode of the signing bonuses of eight NFL players for a specific year. The bonuses in millions of dollars are

18.0, 14.0, 34.5, 10, 11.3, 10, 12.4, 10

Source: *USA TODAY.*

Solution

It is helpful to arrange the data in order although it is not necessary.

10, 10, 10, 11.3, 12.4, 14.0, 18.0, 34.5

Since $10 million occurred 3 times—a frequency larger than any other number—the mode is $10 million.

Example 3–10

Branches of Large Banks

Find the mode for the number of branches that six banks have.

401, 344, 209, 201, 227, 353

Source: SNL Financial.

Solution

Since each value occurs only once, there is no mode.

Note: Do not say that the mode is zero. That would be incorrect, because in some data, such as temperature, zero can be an actual value.

Example 3–11

Licensed Nuclear Reactors

The data show the number of licensed nuclear reactors in the United States for a recent 15-year period. Find the mode.

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

104	104	104	104	104
107	109	109	109	110
109	111	112	111	109

Solution

Since the values 104 and 109 both occur 5 times, the modes are 104 and 109. The data set is said to be bimodal.

The mode for grouped data is the modal class. The **modal class** is the class with the largest frequency.

Example 3–12

Miles Run per Week

Find the modal class for the frequency distribution of miles that 20 runners ran in one week, used in Example 2–7.

Class	Frequency
5.5–10.5	1
10.5–15.5	2
15.5–20.5	3
20.5–25.5	5 ← Modal class
25.5–30.5	4
30.5–35.5	3
35.5–40.5	2

Solution

The modal class is 20.5–25.5, since it has the largest frequency. Sometimes the midpoint of the class is used rather than the boundaries; hence, the mode could also be given as 23 miles per week.

The mode is the only measure of central tendency that can be used in finding the most typical case when the data are nominal or categorical.

Example 3–13 Area Boat Registrations

The data show the number of boats registered for six counties in southwestern Pennsylvania. Find the mode.

Westmoreland	11,008
Butler	9,002
Washington	6,843
Beaver	6,367
Fayette	4,208
Armstrong	3,782

Source: Pennsylvania Fish and Boat Commission.

Solution

Since the category with the highest frequency is Westmoreland, the most typical case is Westmoreland. Hence the mode is 11,008.

An extremely high or extremely low data value in a data set can have a striking effect on the mean of the data set. These extreme values are called *outliers*. This is one reason why when analyzing a frequency distribution, you should be aware of any of these values. For the data set shown in Example 3–14, the mean, median, and mode can be quite different because of extreme values. A method for identifying outliers is given in Section 3–3.

Example 3–14 Salaries of Personnel

A small company consists of the owner, the manager, the salesperson, and two technicians, all of whose annual salaries are listed here. (Assume that this is the entire population.)

Staff	Salary
Owner	$50,000
Manager	20,000
Salesperson	12,000
Technician	9,000
Technician	9,000

Find the mean, median, and mode.

Solution

$$\mu = \frac{\Sigma X}{N} = \frac{50{,}000 + 20{,}000 + 12{,}000 + 9000 + 9000}{5} = \$20{,}000$$

Hence, the mean is $20,000, the median is $12,000, and the mode is $9,000.

In Example 3–14, the mean is much higher than the median or the mode. This is so because the extremely high salary of the owner tends to raise the value of the mean. In this and similar situations, the median should be used as the measure of central tendency.

The Midrange

The *midrange* is a rough estimate of the middle. It is found by adding the lowest and highest values in the data set and dividing by 2. It is a very rough estimate of the average and can be affected by one extremely high or low value.

The **midrange** is defined as the sum of the lowest and highest values in the data set, divided by 2. The symbol MR is used for the midrange.

$$\text{MR} = \frac{\text{lowest value} + \text{highest value}}{2}$$

Example 3–15 Water-Line Breaks

In the last two winter seasons, the city of Brownsville, Minnesota, reported these numbers of water-line breaks per month. Find the midrange.

2, 3, 6, 8, 4, 1

Solution

$$\text{MR} = \frac{1 + 8}{2} = \frac{9}{2} = 4.5$$

Hence, the midrange is 4.5.

If the data set contains one extremely large value or one extremely small value, a higher or lower midrange value will result and may not be a typical description of the middle.

Example 3–16 NFL Signing Bonuses

Find the midrange of data for the NFL signing bonuses in Example 3–9. The bonuses in millions of dollars are

18.0, 14.0, 34.5, 10, 11.3, 10, 12.4, 10

Solution

The smallest bonus is \$10 million and the largest bonus is \$34.5 million.

$$\text{MR} = \frac{10 + 34.5}{2} = \frac{44.5}{2} = \$22.25 \text{ million}$$

Notice that this amount is larger than seven of the eight amounts and is not typical of the average of the bonuses. The reason is that there is one very high bonus, namely, \$34.5 million.

In statistics, several measures can be used for an average. The most common measures are the mean, median, mode, and midrange. Each has its own specific purpose and use. Exercises 39 through 41 show examples of other averages, such as the harmonic mean, the geometric mean, and the quadratic mean. Their applications are limited to specific areas, as shown in the exercises.

The Weighted Mean

Sometimes, you must find the mean of a data set in which not all values are equally represented. Consider the case of finding the average cost of a gallon of gasoline for three taxis. Suppose the drivers buy gasoline at three different service stations at a cost of $3.22, $3.53, and $3.63 per gallon. You might try to find the average by using the formula

$$\overline{X} = \frac{\Sigma X}{n}$$
$$= \frac{3.22 + 3.53 + 3.63}{3} = \frac{10.38}{3} = \$3.46$$

But not all drivers purchased the same number of gallons. Hence, to find the true average cost per gallon, you must take into consideration the number of gallons each driver purchased.

The type of mean that considers an additional factor is called the *weighted mean,* and it is used when the values are not all equally represented.

Interesting Fact

The average American drives about 10,000 miles a year.

Find the **weighted mean** of a variable X by multiplying each value by its corresponding weight and dividing the sum of the products by the sum of the weights.

$$\overline{X} = \frac{w_1X_1 + w_2X_2 + \cdots + w_nX_n}{w_1 + w_2 + \cdots + w_n} = \frac{\Sigma wX}{\Sigma w}$$

where $w_1, w_2, \ldots, w_n$ are the weights and $X_1, X_2, \ldots, X_n$ are the values.

Example 3–17 shows how the weighted mean is used to compute a grade point average. Since courses vary in their credit value, the number of credits must be used as weights.

Example 3–17

Grade Point Average

A student received an A in English Composition I (3 credits), a C in Introduction to Psychology (3 credits), a B in Biology I (4 credits), and a D in Physical Education (2 credits). Assuming A = 4 grade points, B = 3 grade points, C = 2 grade points, D = 1 grade point, and F = 0 grade points, find the student's grade point average.

Solution

Course	Credits (w)	Grade (X)
English Composition I	3	A (4 points)
Introduction to Psychology	3	C (2 points)
Biology I	4	B (3 points)
Physical Education	2	D (1 point)

$$\overline{X} = \frac{\Sigma wX}{\Sigma w} = \frac{3 \cdot 4 + 3 \cdot 2 + 4 \cdot 3 + 2 \cdot 1}{3 + 3 + 4 + 2} = \frac{32}{12} = 2.7$$

The grade point average is 2.7.

Table 3–1 summarizes the measures of central tendency.

Unusual Stat

Of people in the United States, 45% live within 15 minutes of their best friend.

Table 3–1 Summary of Measures of Central Tendency

Measure	Definition	Symbol(s)
Mean	Sum of values, divided by total number of values	μ, $\overline{X}$
Median	Middle point in data set that has been ordered	MD
Mode	Most frequent data value	None
Midrange	Lowest value plus highest value, divided by 2	MR

Researchers and statisticians must know which measure of central tendency is being used and when to use each measure of central tendency. The properties and uses of the four measures of central tendency are summarized next.

Properties and Uses of Central Tendency

The Mean

1. The mean is found by using all the values of the data.
2. The mean varies less than the median or mode when samples are taken from the same population and all three measures are computed for these samples.
3. The mean is used in computing other statistics, such as the variance.
4. The mean for the data set is unique and not necessarily one of the data values.
5. The mean cannot be computed for the data in a frequency distribution that has an open-ended class.
6. The mean is affected by extremely high or low values, called outliers, and may not be the appropriate average to use in these situations.

The Median

1. The median is used to find the center or middle value of a data set.
2. The median is used when it is necessary to find out whether the data values fall into the upper half or lower half of the distribution.
3. The median is used for an open-ended distribution.
4. The median is affected less than the mean by extremely high or extremely low values.

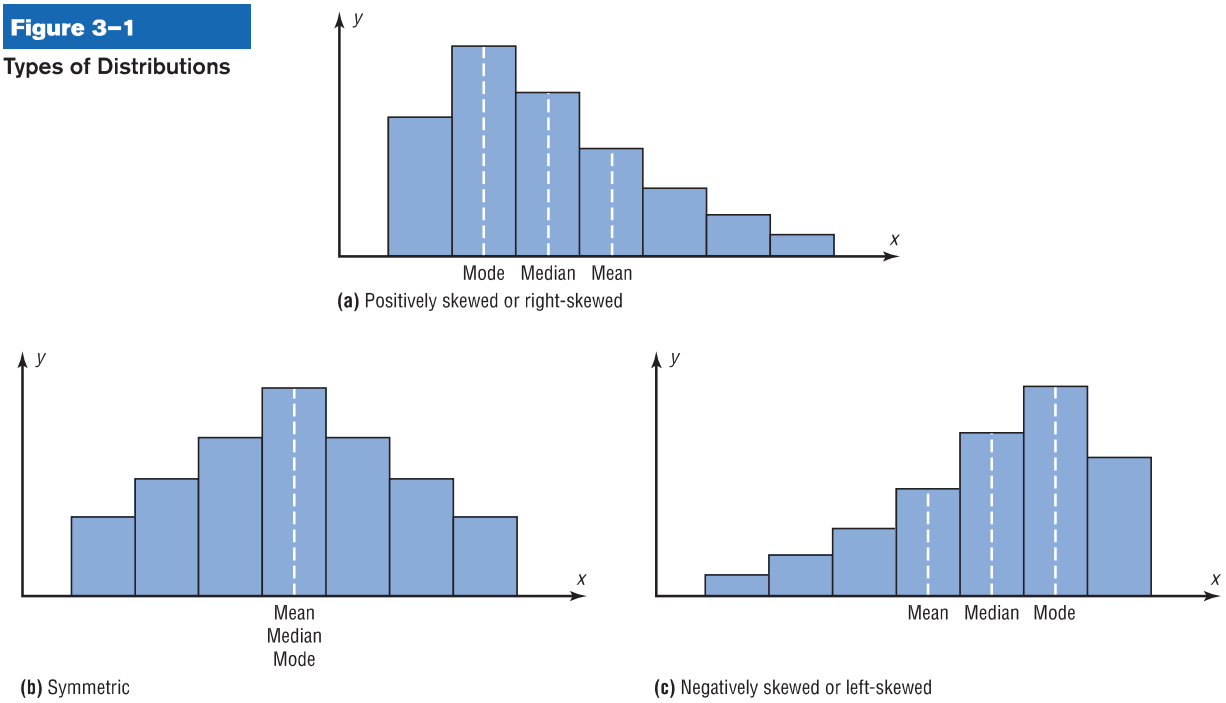
The Mode

1. The mode is used when the most typical case is desired.
2. The mode is the easiest average to compute.
3. The mode can be used when the data are nominal or categorical, such as religious preference, gender, or political affiliation.
4. The mode is not always unique. A data set can have more than one mode, or the mode may not exist for a data set.

The Midrange

1. The midrange is easy to compute.
2. The midrange gives the midpoint.
3. The midrange is affected by extremely high or low values in a data set.

Figure 3–1
Types of Distributions



Distribution Shapes

Frequency distributions can assume many shapes. The three most important shapes are positively skewed, symmetric, and negatively skewed. Figure 3–1 shows histograms of each.

In a **positively skewed** or **right-skewed distribution,** the majority of the data values fall to the left of the mean and cluster at the lower end of the distribution; the "tail" is to the right. Also, the mean is to the right of the median, and the mode is to the left of the median.

For example, if an instructor gave an examination and most of the students did poorly, their scores would tend to cluster on the left side of the distribution. A few high scores would constitute the tail of the distribution, which would be on the right side. Another example of a positively skewed distribution is the incomes of the population of the United States. Most of the incomes cluster about the low end of the distribution; those with high incomes are in the minority and are in the tail at the right of the distribution.

In a **symmetric distribution,** the data values are evenly distributed on both sides of the mean. In addition, when the distribution is unimodal, the mean, median, and mode are the same and are at the center of the distribution. Examples of symmetric distributions are IQ scores and heights of adult males.

When the majority of the data values fall to the right of the mean and cluster at the upper end of the distribution, with the tail to the left, the distribution is said to be **negatively skewed** or **left-skewed.** Also, the mean is to the left of the median, and the mode is to the right of the median. As an example, a negatively skewed distribution results if the majority of students score very high on an instructor's examination. These scores will tend to cluster to the right of the distribution.

When a distribution is extremely skewed, the value of the mean will be pulled toward the tail, but the majority of the data values will be greater than the mean or less than the mean (depending on which way the data are skewed); hence, the median rather than the mean is a more appropriate measure of central tendency. An extremely skewed distribution can also affect other statistics.

A measure of skewness for a distribution is discussed in Exercise 48 in Section 3–2.

Applying the Concepts 3–1

Teacher Salaries

The following data represent salaries (in dollars) from a school district in Greenwood, South Carolina.

10,000	11,000	11,000	12,500	14,300	17,500
18,000	16,600	19,200	21,560	16,400	107,000

1. First, assume you work for the school board in Greenwood and do not wish to raise taxes to increase salaries. Compute the mean, median, and mode, and decide which one would best support your position to not raise salaries.
2. Second, assume you work for the teachers' union and want a raise for the teachers. Use the best measure of central tendency to support your position.
3. Explain how outliers can be used to support one or the other position.
4. If the salaries represented every teacher in the school district, would the averages be parameters or statistics?
5. Which measure of central tendency can be misleading when a data set contains outliers?
6. When you are comparing the measures of central tendency, does the distribution display any skewness? Explain.

See page 180 for the answers.

Exercises 3–1

For Exercises 1 through 9, find (*a*) the mean, (*b*) the median, (*c*) the mode, and (*d*) the midrange.

1. Grade Point Averages The average undergraduate grade point average (GPA) for the 25 top-ranked medical schools is listed below. *a.* 3.724 *b.* 3.73 *c.* 3.74 and 3.70 *d.* 3.715

3.80	3.77	3.70	3.74	3.70
3.86	3.76	3.68	3.67	3.57
3.83	3.70	3.80	3.74	3.67
3.78	3.74	3.73	3.65	3.66
3.75	3.64	3.78	3.73	3.64

Source: *U.S. News & World Report Best Graduate Schools.*

2. Airport Parking The number of short-term parking spaces at 15 airports is shown. *a.* 3174.6 *b.* 1479 *c.* No mode *d.* 5012.5

750	3400	1962	700	203
900	8662	260	1479	5905
9239	690	9822	1131	2516

Source: *USA Today.*

3. High Temperatures The reported high temperatures (in degrees Fahrenheit) for selected world cities on an October day are shown below. Which measure of central tendency do you think best describes these data?

62	72	66	79	83	61	62	85	72	64	74	71
42	38	91	66	77	90	74	63	64	68	42	

Source: www.accuweather.com *a.* 68.1 *b.* 68 *c.* 42, 62, 64, 66, 72, 74 *d.* 64.5

4. Observers in the Frogwatch Program The number of observers in the Frogwatch USA program (a wildlife conservation program dedicated to helping conserve frogs and toads) for the top 10 states with the most observers is 484, 483, 422, 396, 378, 352, 338, 331, 318, and 302. The top 10 states with the most active watchers list these numbers of visits: 634, 464, 406, 267, 219, 194, 191, 150, 130, and 114. Compare the measures of central tendency for these two groups of data.

Source: www.nwf.org/frogwatch

5. Expenditures per Pupil for Selected States The expenditures per pupil for selected states are listed below. Based on these data, what do you think of the claim that the average expenditure per pupil in the United States exceeds $10,000? *a.* 9422.2 *b.* 8988 *c.* 7552, 12,568, 8632 *d.* 9434. Claim seems a little high.

6,300	11,847	8,319	9,344	9,870
10,460	7,491	7,552	12,568	8,632
7,552	12,568	8,632	11,057	10,454
8,109				

Source: *New York Times Almanac.*

6. Earnings of Nonliving Celebrities *Forbes* magazine prints an annual Top-Earning Nonliving Celebrities list (based on royalties and estate earnings). Find the measures of central tendency for these data and comment on the skewness. Figures represent millions of dollars. *a.* 19 *b.* 10 *c.* 7 *d.* 28.5 (Isn't it cool that Albert Einstein is on this list?)

Kurt Cobain	50	Ray Charles	10
Elvis Presley	42	Marilyn Monroe	8
Charles M. Schulz	35	Johnny Cash	8
John Lennon	24	J.R.R. Tolkien	7
Albert Einstein	20	George Harrison	7
Andy Warhol	19	Bob Marley	7
Theodore Geisel (Dr. Seuss)	10		

Source: articles.moneycentral.msn.com

7. Earthquake Strengths Twelve major earthquakes had Richter magnitudes shown here.

7.0, 6.2, 7.7, 8.0, 6.4, 6.2, 7.2, 5.4, 6.4, 6.5, 7.2, 5.4

Which would you consider the best measure of average?

Source: *The Universal Almanac.*

a. 6.63 *b.* 6.45 *c.* 5.4, 6.2, 6.4, 7.2 *d.* 6.7; answers will vary

8. Top-Paid CEOs The data shown are the total compensation (in millions of dollars) for the 50 top-paid CEOs for a recent year. Compare the averages, and state which one you think is the best measure.

17.5	18.0	36.8	31.7	31.7
17.3	24.3	47.7	38.5	17.0
23.7	16.5	25.1	17.4	18.0
37.6	19.7	21.4	28.6	21.6
19.3	20.0	16.9	25.2	19.8
25.0	17.2	20.4	20.1	29.1
19.1	25.2	23.2	25.9	24.0
41.7	24.0	16.8	26.8	31.4
16.9	17.2	24.1	35.2	19.1
22.9	18.2	25.4	35.4	25.5

Source: *USA TODAY.* 24.42; 23.45; 16.9, 17.2, 18, 19.1, 24, 25.2, 31.7; 32.1. It appears that the mean and median are good measures of the average.

9. Garbage Collection The amount of garbage in millions of tons collected over a 16-year period is shown. *a.* 46.78 *b.* 47.65 *c.* None *d.* 44.05

29.7	47.3	32.9	36
48	57.2	53.7	52.8
58.4	55.8	46.1	46.4
37.9	43.5	50.1	52.7

Source: *Environmental Protection Agency.*

10. Foreign Workers The number of foreign workers' certificates for the New England states and the northwestern states is shown. Find the mean, median, and mode for both areas and compare the results.

New England States	Northwest States
6768	1870
3196	622
1112	620
819	23
1019	172
1795	112

Source: Department of Labor.

11. Populations of Selected Cities Populations for towns and cities of 5000 or more (based on the 2004 figures) in the 15XXX zip code area are listed here for two different years. Find the mean, median, mode, and midrange for each set of data. What do your findings suggest?

2004			1990		
11,270	8,825	7,439	13,374	9,200	8,133
8,220	5,132	8,395	9,278	4,768	9,135
5,463	8,174	5,044	6,113	9,656	5,784
8,739	5,282	7,869	9,229	21,923	8,286
6,199	5,307	10,493	10,687	5,319	9,126
10,309	14,925	8,397	11,221	15,174	9,901
9,964	14,849	5,094	10,823	15,864	5,445
14,340	5,707	6,672	14,292	5,748	6,961

Source: *World Almanac.*

For Exercises 12 through 21, find the (*a*) mean and (*b*) modal class.

12. Executive Bonuses A random sample of bonuses (in millions of dollars) paid by large companies to their executives is shown. These data will be used for Exercise 18 in Section 3–2. *a.* 5 *b.* 3.5–6.5

Class boundaries	Frequency
0.5–3.5	11
3.5–6.5	12
6.5–9.5	4
9.5–12.5	2
12.5–15.5	1

13. Hourly Compensation for Production Workers The hourly compensation costs (in U.S. dollars) for production workers in selected countries are represented below.

Class	Frequency
2.48–7.48	7
7.49–12.49	3
12.50–17.50	1
17.51–22.51	7
22.52–27.52	5
27.53–32.53	5

a. 17.68 *b.* 2.48–7.48 and 17.51–22.51. Group mean is less.

Compare the mean of these grouped data to the U.S. mean of $21.97.

Source: *New York Times Almanac.*

14. Automobile Fuel Efficiency Thirty automobiles were tested for fuel efficiency (in miles per gallon). This frequency distribution was obtained. (The data in this exercise will be used in Exercise 20 in Section 3–2.) *a.* 19.7 *b.* 17.5–22.5

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

15. Percentage of Foreign-Born People The percentage of foreign-born population for each of the 50 states is represented below. Do you think the mean is the best average for this set of data? Explain. *a.* 6.5 *b.* 0.8–4.4. Probably not—data are "top heavy."

Percentage	Frequency
0.8–4.4	26
4.5–8.1	11
8.2–11.8	4
11.9–15.5	5
15.6–19.2	2
19.3–22.9	1
23.0–26.6	1

Source: *World Almanac.*

16. Find the mean and modal class for each set of data in Exercises 8 and 18 in Section 2–2. Is the average about the same for both sets of data?

17. Percentage of College-Educated Population over 25 Below are the percentages of the population over 25 years of age who have completed 4 years of college or more for the 50 states and the District of Columbia. Find the mean and modal class. *a.* 26.7 *b.* 24.2–28.6

Percentage	Frequency
15.2–19.6	3
19.7–24.1	15
24.2–28.6	19
28.7–33.1	6
33.2–37.6	7
37.7–42.1	0
42.2–46.6	1

Source: *New York Times Almanac.*

18. Net Worth of Corporations These data represent the net worth (in millions of dollars) of 45 national corporations. *a.* 42.9 *b.* 32–42

Class limits	Frequency
10–20	2
21–31	8
32–42	15
43–53	7
54–64	10
65–75	3

19. Specialty Coffee Shops A random sample of 30 states shows the number of specialty coffee shops for a specific company. *a.* 34.1 *b.* 0.5–19.5

Class boundaries	Frequency
0.5–19.5	12
19.5–38.5	7
38.5–57.5	5
57.5–76.5	3
76.5–95.5	3

20. Commissions Earned This frequency distribution represents the commission earned (in dollars) by 100 salespeople employed at several branches of a large chain store. *a.* 180.3 *b.* 177–185

Class limits	Frequency
150–158	5
159–167	16
168–176	20
177–185	21
186–194	20
195–203	15
204–212	3

21. Copier Service Calls This frequency distribution represents the data obtained from a sample of 75 copying machine service technicians. The values represent the days between service calls for various copying machines. *a.* 23.7 *b.* 21.5–24.5

Class boundaries	Frequency
15.5–18.5	14
18.5–21.5	12
21.5–24.5	18
24.5–27.5	10
27.5–30.5	15
30.5–33.5	6

22. Use the data from Exercise 14 in Section 2–1 and find the mean and modal class. *a.* 14.6 *b.* 0–10

23. Find the mean and modal class for the data in Exercise 13 in Section 2–1. 44.8; 40.5–47.5

24. Use the data from Exercise 3 in Section 2–2 and find the mean and modal class. *a.* 64.4 *b.* 3–45 and 46–88

25. Enrollments for Selected Independent Religiously Controlled 4-Year Colleges Listed below are the enrollments for selected independent religiously controlled 4-year colleges that offer bachelor's degrees only. Construct a grouped frequency distribution with six classes and find the mean and modal class. *a.* 1804.6 *b.* 1013–1345

1013 1867 1268 1666 2309 1231 3005 2895 2166 1136
1532 1461 1750 1069 1723 1827 1155 1714 2391 2155
1412 1688 2471 1759 3008 2511 2577 1082 1067 1062
1319 1037 2400

Source: *World Almanac.*

26. Find the weighted mean price of three models of automobiles sold. The number and price of each model sold are shown in this list. $9866.67

Model	Number	Price
A	8	$10,000
B	10	12,000
C	12	8,000

27. Fat Grams Using the weighted mean, find the average number of grams of fat per ounce of meat or fish that a person would consume over a 5-day period if he ate these:

Meat or fish	Fat (g/oz)
3 oz fried shrimp	3.33
3 oz veal cutlet (broiled)	3.00
2 oz roast beef (lean)	2.50
2.5 oz fried chicken drumstick	4.40
4 oz tuna (canned in oil)	1.75

Source: *The World Almanac and Book of Facts.* 2.896

28. Diet Cola Preference A recent survey of a new diet cola reported the following percentages of people who liked the taste. Find the weighted mean of the percentages. 35.4%

Area	% Favored	Number surveyed
1	40	1000
2	30	3000
3	50	800

29. Costs of Helicopters The costs of three models of helicopters are shown here. Find the weighted mean of the costs of the models. $545,666.67

Model	Number sold	Cost
Sunscraper	9	$427,000
Skycoaster	6	365,000
High-flyer	12	725,000

30. Final Grade An instructor grades exams, 20%; term paper, 30%; final exam, 50%. A student had grades of 83, 72, and 90, respectively, for exams, term paper, and final exam. Find the student's final average. Use the weighted mean. 83.2

31. Final Grade Another instructor gives four 1-hour exams and one final exam, which counts as two 1-hour exams. Find a student's grade if she received 62, 83, 97, and 90 on the 1-hour exams and 82 on the final exam. 82.7

32. For these situations, state which measure of central tendency—mean, median, or mode—should be used.

a. The most typical case is desired. Mode
b. The distribution is open-ended. Median
c. There is an extreme value in the data set. Median
d. The data are categorical. Mode
e. Further statistical computations will be needed. Mean
f. The values are to be divided into two approximately equal groups, one group containing the larger values and one containing the smaller values. Median

33. Describe which measure of central tendency—mean, median, or mode—was probably used in each situation.

a. One-half of the factory workers make more than $5.37 per hour, and one-half make less than $5.37 per hour. Median
b. The average number of children per family in the Plaza Heights Complex is 1.8. Mean
c. Most people prefer red convertibles over any other color. Mode
d. The average person cuts the lawn once a week. Mode
e. The most common fear today is fear of speaking in public. Mode
f. The average age of college professors is 42.3 years. Mean

34. What types of symbols are used to represent sample statistics? Give an example. What types of symbols are used to represent population parameters? Give an example. Roman letters, $\bar{X}$; Greek letters, μ

35. A local fast-food company claims that the average salary of its employees is $13.23 per hour. An employee states that most employees make minimum wage. If both are being truthful, how could both be correct? Both could be true since one may be using the mean for the average salary and the other may be using the mode for the average.

Extending the Concepts

36. If the mean of five values is 64, find the sum of the values. 320

37. If the mean of five values is 8.2 and four of the values are 6, 10, 7, and 12, find the fifth value. 6

38. Find the mean of 10, 20, 30, 40, and 50.

a. Add 10 to each value and find the mean. 40
b. Subtract 10 from each value and find the mean. 20
c. Multiply each value by 10 and find the mean. 300
d. Divide each value by 10 and find the mean. 3
e. Make a general statement about each situation. The results will be the same as if you add, subtract, multiply, and divide the mean by 10.

39. The *harmonic mean* (HM) is defined as the number of values divided by the sum of the reciprocals of each value. The formula is

$$\text{HM} = \frac{n}{\Sigma(1/X)}$$

For example, the harmonic mean of 1, 4, 5, and 2 is

$$\text{HM} = \frac{4}{1/1 + 1/4 + 1/5 + 1/2} = 2.05$$

This mean is useful for finding the average speed. Suppose a person drove 100 miles at 40 miles per hour and returned driving 50 miles per hour. The average miles per hour is *not* 45 miles per hour, which is found by adding 40 and 50 and dividing by 2. The average is found as shown.

Since

$$\text{Time} = \text{distance} \div \text{rate}$$

then

$$\text{Time 1} = \frac{100}{40} = 2.5 \text{ hours to make the trip}$$

$$\text{Time 2} = \frac{100}{50} = 2 \text{ hours to return}$$

Hence, the total time is 4.5 hours, and the total miles driven are 200. Now, the average speed is

$$\text{Rate} = \frac{\text{distance}}{\text{time}} = \frac{200}{4.5} = 44.44 \text{ miles per hour}$$

This value can also be found by using the harmonic mean formula

$$\text{HM} = \frac{2}{1/40 + 1/50} = 44.44$$

Using the harmonic mean, find each of these.

a. A salesperson drives 300 miles round trip at 30 miles per hour going to Chicago and 45 miles per hour returning home. Find the average miles per hour. 36 mph

b. A bus driver drives the 50 miles to West Chester at 40 miles per hour and returns driving 25 miles per hour. Find the average miles per hour. 30.77 mph

c. A carpenter buys \$500 worth of nails at \$50 per pound and \$500 worth of nails at \$10 per pound. Find the average cost of 1 pound of nails. \$16.67

40. The *geometric mean* (GM) is defined as the nth root of the product of n values. The formula is

$$\text{GM} = \sqrt[n]{(X_1)(X_2)(X_3)\cdots(X_n)}$$

The geometric mean of 4 and 16 is

$$\text{GM} = \sqrt{(4)(16)} = \sqrt{64} = 8$$

The geometric mean of 1, 3, and 9 is

$$\text{GM} = \sqrt[3]{(1)(3)(9)} = \sqrt[3]{27} = 3$$

The geometric mean is useful in finding the average of percentages, ratios, indexes, or growth rates. For example, if a person receives a 20% raise after 1 year of service and a 10% raise after the second year of service, the average percentage raise per year is not 15 but 14.89%, as shown.

$$\text{GM} = \sqrt{(1.2)(1.1)} = 1.1489$$

or

$$\text{GM} = \sqrt{(120)(110)} = 114.89\%$$

His salary is 120% at the end of the first year and 110% at the end of the second year. This is equivalent to an average of 14.89%, since 114.89% − 100% = 14.89%.

This answer can also be shown by assuming that the person makes \$10,000 to start and receives two raises of 20 and 10%.

$$\text{Raise 1} = 10{,}000 \cdot 20\% = \$2000$$
$$\text{Raise 2} = 12{,}000 \cdot 10\% = \$1200$$

His total salary raise is \$3200. This total is equivalent to

$$\begin{aligned} \$10{,}000 \cdot 14.89\% &= \$1489.00 \\ \$11{,}489 \cdot 14.89\% &= \quad 1710.71 \\ &\overline{\qquad\qquad} \\ &\$3199.71 \approx \$3200 \end{aligned}$$

Find the geometric mean of each of these.

a. The growth rates of the Living Life Insurance Corporation for the past 3 years were 35, 24, and 18%. 25.5%

b. A person received these percentage raises in salary over a 4-year period: 8, 6, 4, and 5%. 5.7%

c. A stock increased each year for 5 years at these percentages: 10, 8, 12, 9, and 3%. 8.4%

d. The price increases, in percentages, for the cost of food in a specific geographic region for the past 3 years were 1, 3, and 5.5%. 3.2%

41. A useful mean in the physical sciences (such as voltage) is the *quadratic mean* (QM), which is found by taking the square root of the average of the squares of each value. The formula is

$$\text{QM} = \sqrt{\frac{\Sigma X^2}{n}}$$

The quadratic mean of 3, 5, 6, and 10 is

$$\text{QM} = \sqrt{\frac{3^2 + 5^2 + 6^2 + 10^2}{4}} = \sqrt{42.5} = 6.52$$

Find the quadratic mean of 8, 6, 3, 5, and 4. 5.48

42. An approximate median can be found for data that have been grouped into a frequency distribution. First it is necessary to find the median class. This is the class that contains the median value. That is the $n/2$ data value. Then it is assumed that the data values are evenly distributed throughout the median class. The formula is

$$\text{MD} = \frac{n/2 - \text{cf}}{f}(w) + L_m$$

where n = sum of frequencies
cf = cumulative frequency of class immediately preceding the median class
w = width of median class
f = frequency of median class
L_m = lower boundary of median class

Using this formula, find the median for data in the frequency distribution of Exercise 15. 4.31