CHAPTER 4 Measures of Central Tendency

Criminal justice and criminology researchers and practitioners are often interested in averages. Averages offer information about the centers or middles of distribution. They indicate where data points tend to cluster. This is important to know. Consider the following questions that might be of interest to a researcher or practitioner:

1. What is the most common level of educational attainment among police officers?
2. How does the median income for people living in a socioeconomically disadvantaged area of a certain city compare to that for all people in the city?
3. What is the average violent crime rate across all cities and towns in a particular state?
4. Do prison inmates, in general, have a lower average reading level compared to the general population?

All of these questions make some reference to an average, a middle point, or, to use a more technical term, a **measure of central tendency**. Measures of central tendency offer information about where the bulk of the scores in a particular data set are located. A person who is computing a measure of central tendency is, in essence, asking, "Where is the middle?"

Averages offer information about the normal or typical person, object, or place in a sample. A group of people with an average age of 22, for instance, probably looks different from a group averaging 70 years of age. Group averages help us predict the score for any individual within that group. Suppose in two samples of people, the only information you have is that one group's average weight is 145 pounds and that the other's is 200 pounds. If someone asked you, "How much does an individual person in the first group weigh?" your response would be, "About 145 pounds." If you were asked, "Who weighs more, a person randomly selected from the first group or from the second group?" you would respond that the person from the

Learning Objectives

- Define the three types of data distribution shapes.
- Identify the shape of a distribution based on a comparison of the mean and median.
- Describe the mode, the median, and the mean.
- Explain which level(s) of measurement each measure of central tendency can be used with.
- Identify, locate, or calculate the mode, the median, and the mean for a variety of variables and variable types.
- Explain the difference between the two mean formulas and correctly use each one on the appropriate types of data.
- Explain deviation scores and their relationship to the mean.
- Use SPSS to produce the mode, the median, and the mean.

Measures of central tendency: Descriptive statistics that offer information about where the scores in a particular data set tend to cluster. Examples include the mode, the median, and the mean.

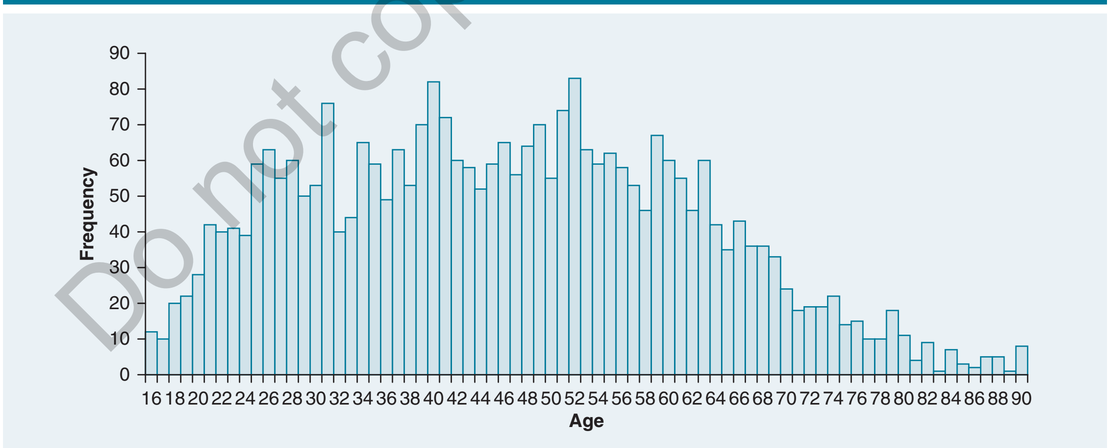
Normal distribution: A set of scores that clusters in the center and tapers off to the left (negative) and right (positive) sides of the number line.

second group is probably the heavier of the two. Of course, you do not know for sure that you are correct; there might be people in the first group who are heavier than some people in the second group. The average, nonetheless, gives you predictive capability. It allows you to draw general conclusions and to form a basic level of understanding about a set of objects, places, or people.

Measures of central tendency speak to the matter of distribution shape. Data distributions come in many different shapes and sizes. Figure 4.1 contains data from the Police–Public Contact Survey (PPCS; see Data Sources 2.1) showing the ages of non-Hispanic respondents who reported having called the police for help within the past 24 months. This is the same variable used in the frequency polygon shown in Figure 3.10 in the previous chapter. The shape this variable assumes is called a **normal distribution**. The normal curve represents an even distribution of scores. The most frequently occurring values are in the middle of the curve, and frequencies drop off as one traces the number line to the left or right. Normal distributions are ideal in research because the average is truly the best predictor of the scores for each case in the sample, since the scores cluster around that value.

Standing in contrast to normal curves are skewed distributions. Skew can be either positive or negative. The distribution in Figure 4.2 contains the data from the Census of Jails (COJ; see Data Sources 3.1) showing the number of inmate-on-staff assaults each jail reported experiencing in the past year. The distribution in Figure 4.2 manifests what is called a **positive skew**. Positively skewed data cluster on the left-hand side of the distribution, with extreme values in the right-hand portion that pull the tail out toward the positive side of the number line. Positively skewed data are common in criminal justice and criminology research.

Figure 4.1 Ages of Non-Hispanic Respondents Who Called the Police in the Past 24 Months



Learning Check 4.1

Skew type (positive versus negative) is determined by the location of the elongated tail of a skewed distribution. Positively skewed distributions are those in which the tail extends toward the positive side of the number line; likewise, negative skew is signaled by a tail extending toward negative infinity. Set aside your book and draw one of each type of distribution from memory.

Positive skew: A clustering of scores in the left-hand side of a distribution with some relatively large scores that pull the tail toward the positive side of the number line.

Figure 4.2 Number of Inmate Assaults on Jail Staff

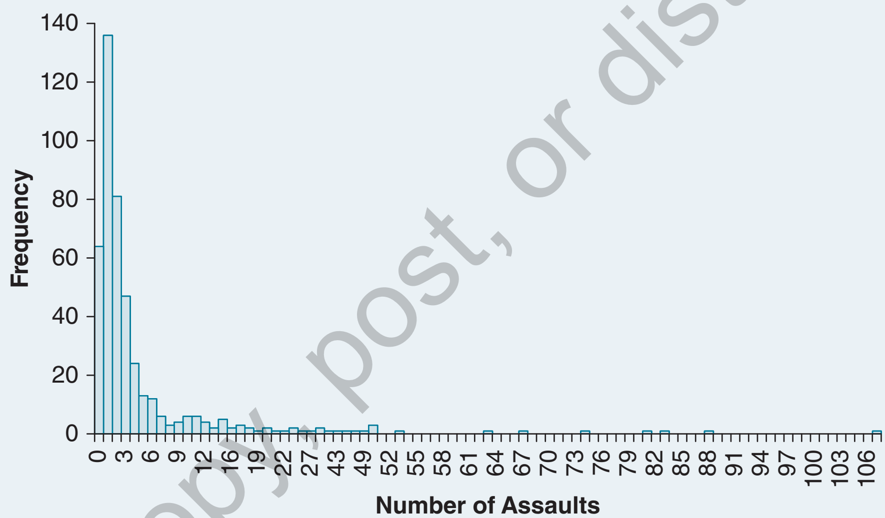
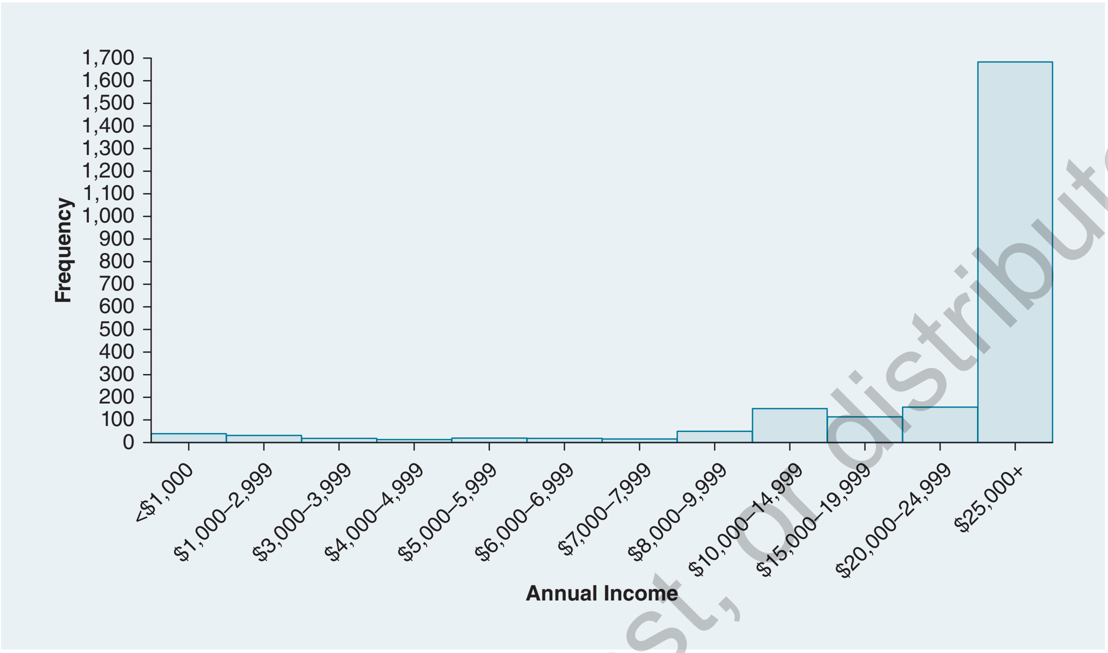


Figure 4.3 shows 2014 General Social Survey (GSS; see Data Sources 2.2) respondents' annual family incomes. This distribution has a **negative skew**: Scores are sparse on the left-hand side, and they increase in frequency on the right side of the distribution.

Negative skew: A clustering of scores in the right-hand side of a distribution with some relatively small scores that pull the tail toward the negative side of the number line.

Knowing whether a given distribution of data points is normal or skewed is vital in criminal justice and criminology research. The average is an excellent predictor of individual scores when the curve is normal. When a distribution departs from normality, however, the average becomes less useful and, in extreme cases, can be misleading. For example, the mean number of inmate-on-staff assaults is 5.5 per jail, but you can see in Figure 4.2 that the vast majority of jails had four or fewer assaults; in fact, a full two-thirds experienced just two, one, or even zero assaults. A statement such as "Jails had an average of 5.5 inmate-on-staff assaults in 2013" would be technically correct,

Figure 4.3 GSS Respondents' Annual Family Incomes



but would be very misleading because the typical jail has fewer incidents, and many have a great deal more than that as well. Because this distribution is so skewed, the mean loses its usefulness as a description of the middle of the data. Distribution shape plays a key role in more-complicated statistical analyses (more on this in Parts II and III) and is important in a descriptive sense so that information conveyed to academic, practitioner, or lay audiences is fully accurate and transparent. You must always know where the middle of your data set is located; measures of central tendency give you that information.

The Mode

Mode: The most frequently occurring category or value in a set of scores.

The **mode** is the simplest of the three measures of central tendency covered in this chapter. It requires no mathematical computations and can be employed with any level of measurement. It is the only measure of central tendency available for use with nominal data. The mode is simply the most frequently occurring category or value. Table 4.1 contains data from the 2011 PPCS (Data Sources 2.1). Interviewers asked PPCS respondents whether they had been stopped by police while driving a vehicle. The people who answered yes were then asked to report the reason for that stop. This is a nominal-level variable. Table 4.1 presents the distribution of responses that participants gave for their stop. The mode is *speeding* because that is the stop reason that occurs most frequently (i.e., 2,040 people said that this is the violation for which they were pulled over by police).

A frequency bar graph of the same data is shown in Figure 4.4. The mode is easily identifiable as the category accompanied by the highest bar.

The mode can also be used with continuous variables. Instead of identifying the most frequently occurring category as with nominal or ordinal data, you will identify the most common value. Figure 4.5 shows a frequency histogram for the variable from the PPCS that asks respondents how many face-to-face contacts they had with the police in the past 12 months. The sample has been narrowed to include only female respondents who were 21 or younger at the time of the survey. Can you identify the modal number of contacts? If you answered "1," you are correct!

Table 4.1 Among Stopped Drivers, Reason for the Stop

Stop Reason	*Frequency*
Speeding	2,040
Vehicle Defect	599
Record Check	381
Roadside Check	66
Seatbelt Violation	202
Illegal Turn or Lane Change	257
Stop Sign or Light Violation	275
Cellphone Usage	76
Other	283
Total	$N = 4{,}179$

Learning Check 4.2

Flip back to Table 3.7 in the previous chapter and identify the modal jail size. (Hint: Use the row totals.)

Figure 4.4 Among Stopped Drivers, Reason for the Stop

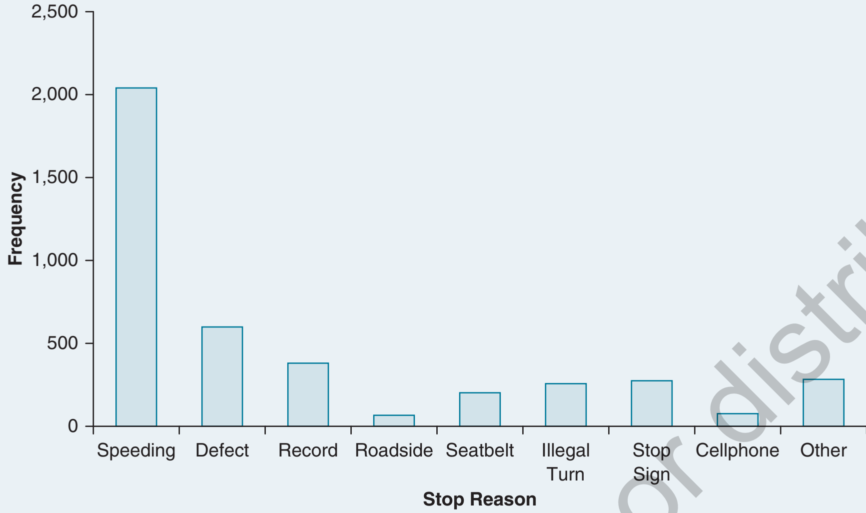
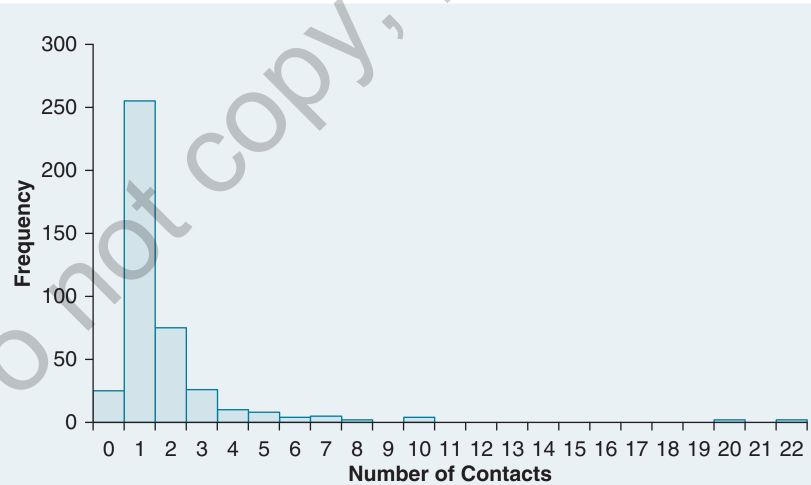


Figure 4.5 Number of Police Contacts in Past 12 Months Among Females Age 21 and Younger



Research Example 4.1

Are People Convicted of Homicide More Violent in Prison Than People Convicted of Other Types of Offenses?

Sorensen and Cunningham (2010) analyzed the institutional conduct records of all inmates incarcerated in Florida state correctional facilities in 2003, along with the records of inmates who entered prison that same year. They divided inmates into three groups. The stock population consisted of all people incarcerated in Florida prisons during 2003, regardless of the year they were admitted into prison. The new persons admitted into prison during 2002 and serving all of 2003 composed the admissions cohort. The close custody group was a subset of the admissions cohort and was made of the inmates who were considered to be especially high threats to institutional security. The table below contains descriptive information about the three samples. For each group, can you identify the modal custody level and modal conviction offense type? (Hint: These are column percentages.) Visit **edge .sagepub.com/gau3e** to view the full article and see the results of this study.

Demographics	*Stock Population (N = 51,527)*	*Admissions Cohort (N = 14,088)*	*Close Custody Sample (N = 4,113)*
Custody level (%)			
Community	6.0	10.4	0.0
Minimum	17.2	23.7	0.0
Medium	27.2	33.8	0.0
Close	48.9	32.0	100.0
Death row	0.6	0.1	0.0
Conviction offense type (%)			
Homicide	18.6	5.9	10.9
Other violent	39.6	33.8	46.2
Property	21.4	26.7	21.0
Drugs	14.7	24.0	14.7
Public order/weapons	5.7	9.6	7.2

Source: Adapted from Table 1 in Sorensen and Cunningham (2010).

(Continued)

(Continued)

Do Latino Drug Traffickers' National Origin and Immigration Status Affect the Sentences They Receive?

The United States is experiencing significant demographic shifts, one of the most notable being the steady increase in the percentage of the population that is Latino. Immigration (both legal and illegal) is a significant contributor to this trend. The influx has inspired many native-born U.S. citizens to react with a get-tough, law enforcement–oriented mind-set toward immigrants who commit crimes in this country. Mexican immigrants, in particular, are frequently seen as a threat, and this belief could translate into harsher treatment for those immigrants who commit crimes. Logue (2017) examined whether Latino drug traffickers' countries of origin and their immigration status impacted the severity of the sentences they receive in federal court. One of the variables she included in her analysis of sentencing outcomes was the type of drug that defendants were alleged to have trafficked. The table shows the breakdown of drug types across defendants of different origins and immigration statuses. Can you identify the modal drug type for each of the four defendant groups? (Hint: These are column percentages.) Visit **edge.sagepub.com/gau3e** to view the full article and see the results of this study.

	Mexican, Legal	*Mexican, Undocumented*	*Non-Mexican, Legal*	*Non-Mexican, Undocumented*
Drug Type (%)				
Powder cocaine	27.5	21.5	44.1	53.6
Crack cocaine	0.4	0.7	3.9	4.3
Heroin	3.7	4.5	28.1	25.9
Marijuana	48.9	49.4	15.4	7.5
Methamphetamine	19.2	23.8	4.5	6.6
Other	0.2	0.1	4.0	2.1

Source: Adapted from Table 1 in Logue (2017).

Learning Check 4.3

Remember that the mode is the category or value that *occurs* the most frequently—it is *not* the frequency itself. Check your ability to tell the difference between the mode and its frequency by looking back at Figure 4.2. Identify the modal number of assaults and the approximate frequency for this number.

The previous two examples highlight the relative usefulness of the mode for categorical data as opposed to continuous data. This measure of central tendency can be informative for the former but is usually not all that interesting or useful for the latter. This is because there are other, more-sophisticated measures that can be calculated with continuous data.

The strengths of the mode include the fact that this measure is simple to identify and understand. It also, as mentioned previously, is the only measure of central tendency that can be used with nominal variables. The mode's major weakness is actually the flipside of its primary strength: Its simplicity means that it is usually too superficial to be of much use. It accounts for only one category or value in the data set and ignores the rest. It also cannot be used in more-complex computations, which greatly limits its usefulness in statistics. The mode, then, can be an informative measure for nominal variables (and sometimes ordinal, as well) and is useful for audiences who are not schooled in statistics, but its utility is restricted, especially with continuous variables.

The Median

The **median** (*Md*) is another measure of central tendency. The median can be used with continuous and ordinal data; it cannot be used with nominal data, however, because it requires that the variable under examination be rank orderable, which nominal variables are not.

Median: The score that cuts a distribution in half such that 50% of the scores are above that value and 50% are below it.

The median is the value that splits a data set exactly in half such that 50% of the data points are below it and 50% are above it. For this reason, the median is also sometimes called the *50th percentile*. The median is a positional measure, which means that it is not so much calculated as it is located. Finding the median is a three-step process. First, the categories or scores need to be rank ordered. Second, the median position (*MP*) can be computed using the formula

$$MP = \frac{N+1}{2} \qquad \text{Formula 4(1)}$$

where

N = total sample size

The median position tells you where the median is located within the ranked data set. The third step is to use the median position to identify the median. When N is odd, the median will be a value in the data set. When N is even, the median will have to be computed by averaging two values.

Let us figure out the median violent crime rate among the five Colorado cities listed in Table 4.2. The variable *violent crime rate* is continuous (specifically, ratio level), and the median is therefore an applicable measure of central tendency. The numbers in Table 4.2 are derived from the 2015 Uniform Crime Reports (UCR; see Data Sources 1.1).

Table 4.2 Violent Crime Rates in Five Colorado Cities

City	Violent Crimes per 10,000 Residents
Aspen	8.77
Colorado Springs	43.83
Denver	67.39
Woodland Park	15.30
Steamboat Springs	22.74
$N = 5$	

The first step is to rank the rates in either ascending or descending order. Ranked in ascending order, they look like this:

$$8.77^1 \quad 15.30^2 \quad 22.74^3 \quad 43.83^4 \quad 67.39^5$$

Superscripts have been inserted in the ranked list to help emphasize the median's nature as a positional measure that is dependent on the *location* of data points rather than these points' actual values. The superscripts represent each number's position in the data set now that the values have been rank ordered.

Second, the formula for the median position will tell you where to look to find the median. Here,

$$MP = \frac{5+1}{2} = \frac{6}{2} = 3$$

This means that the median is in position 3. Remember that the *MP* is *not* the median; rather, it is a "map" that tells you where to look to find the median.

Finally, use the *MP* to identify the median. Since the median is in position 3, we can determine that $Md = 22.74$. This group of five Colorado cities has a median violent crime rate of 22.74 per 10,000.

In this example, the sample had five cases. When there is an even number of cases, finding the median is slightly more complex and requires averaging two numbers together. To demonstrate this, we will use 2015 property crime rates in six North Dakota cities (Table 4.3).

First, rank the values: $38.50^1 \quad 130.37^2 \quad 149.67^3 \quad 239.09^4 \quad 297.24^5 \quad 325.03^6$.

Second, find the median position:

$$MP = \frac{6+1}{2} = \frac{7}{2} = 3.5$$

Table 4.3 Property Crime Rates in Six North Dakota Cities

City	*Property Crimes per 10,000 Residents*
Bismarck	130.37
Fargo	297.24
Burlington	38.50
Minot	325.03
Valley City	239.09
Lisbon	149.67
$N = 6$	

Notice that the *MP* has a decimal this time—this is what happens when the number of cases is even rather than odd. What this means is that the median is halfway between positions 3 and 4, so finding *Md* requires averaging these two numbers. The median is

$$Md = \frac{149.67 + 239.09}{2} = \frac{388.76}{2} = 194.38$$

This sample of six North Dakota cities has a median property crime rate of 194.38 per 10,000 residents. Of these cities, 50% have rates that are lower than this, and 50% have rates that are higher.

For another example of locating the median in an even-numbered sample size, we will use state-level homicide rates (per 100,000 population) from the 2015 UCR. Table 4.4 shows the data. To increase the complexity a bit, we will use eight states.

Table 4.4 Homicide Rates in Eight States

State	*Homicides per 100,000 Residents*
Alabama	7.16
Arkansas	6.08
Illinois	5.79
Hawai'i	1.33
Idaho	1.93
Indiana	5.63
Kansas	4.40
Massachusetts	1.88
$N = 8$	

Following the three steps, we first rank order the values:

$$1.33^1 \quad 1.88^2 \quad 1.93^3 \quad 4.40^4 \quad 5.63^5 \quad 5.79^6 \quad 6.08^7 \quad 7.16^8$$

Next, we use the *MP* formula to locate the position of the median in the data:

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

The *MP* of 4.5 tells us the median is halfway between the numbers located in positions 4 and 5, so we take these two numbers and find their average. The number in position 4 is 4.40 and the number in position 5 is 5.63. Their average is

$$Md = \frac{4.40+5.63}{2} = \frac{10.03}{2} = 5.02$$

The median homicide rate (per 100,000) in this sample of eight states is 5.02.

Learning Check 4.4

Add the homicide rate for the state of Maryland (8.59) to the sample of states in Table 4.4 and locate the median using the three steps. How much did the median change with the inclusion of this state?

Medians can also be found in ordinal-level variables, though the median of an ordinal variable is less precise than that of a continuous variable because the former is a category rather than a number. To demonstrate, we will use an item from the PPCS that measures the driving frequency among respondents who said that they had been pulled over by police for a traffic stop within the past 6 months, either as a driver or as a passenger in the stopped vehicle. Table 4.5 displays the data.

The first two steps to finding the *Md* of ordinal data mirror those for continuous data. First, the median position must be calculated using Formula 4(1). For ordinal data, the total sample size (*N*) is found by summing the frequencies. In Table 4.5, N = 1,544 and so the *MP* is calculated as

$$MP = \frac{1{,}544+1}{2} = \frac{1{,}545}{2} = 772.5$$

The median position in this case is a person—the person who is in position 772.5 is the median.

The second step involves identifying the category in which the *MP* is located. Instead of ranking the categories according to frequencies as we did with continuous data, we are now going to arrange them in either ascending or descending order according to the categories themselves. In other words, the internal ranking system of the categories themselves is used to structure the sequence of the list. In Table 4.5, the categories are arranged from the most-frequent driving habits (*Every Day or Almost Every Day*) to the least-frequent ones (*Never*). As such, the categories are already in descending order and do not need to be rearranged.

Table 4.5 Driving Frequency of PPCS Respondents Who Experienced Traffic Stops

How often do you drive?	f
Every Day or Almost Every Day	751
A Few Days a Week	217
A Few Days a Month	82
A Few Times a Year	37
Never	457
	N = 1,544

Next, add cumulative frequencies of the rank-ordered categories until the sum meets or exceeds the *MP*. Table 4.6 illustrates this process. Here, 751 + 217 = 968, so the median is located in the *A Few Days a Week* category. In other words, if you lined up all 1,544 people in this sample in the order in which they answered the question, labeling the people in each group accordingly, and then counted them until you reached 772.5, the person in that position would be standing in the *A Few Days a Week* group. Therefore, we now know that half the people in this sample drive a few days a week or more, and half drive a few days a week or less.

Table 4.6 Driving Frequency of PPCS Respondents Who Experienced Traffic Stops

How often do you drive?	f	cf
Every Day or Almost Every Day	751	751
A Few Days a Week	217	968
A Few Days a Month	82	1,050
A Few Times a Year	37	1,087
Never	457	1,544
	N = 1,544	

Note how much less informative the median is for ordinal variables as compared to continuous ones. For the crime rates in Tables 4.2 through 4.4, we were able to identify the specific, numerical median; for the driving-frequency variable in Table 4.5, we are able to say only that the median case is contained within the *A Few Days a Week* category. This is a rough estimate that paints a limited picture.

Learning Check 4.5

Rearrange the frequencies from Table 4.5 so that they are in descending order of driving frequency, rather than in ascending order as was the case in the demonstration. Complete the cumulative-frequencies exercise to locate the median. What is your conclusion?

The median has advantages and disadvantages. Its advantages are that it uses more information than the mode does, so it offers a more-descriptive, more-informative picture of the data. It can be used with ordinal variables, which is advantageous because, as we will see, the mean cannot be.

A key advantage of the median is that it is not sensitive to extreme values or outliers. To understand this concept, revisit Table 4.3 and replace Minot's rate of 325.03 with 600.00; then relocate the median. It did not change, despite a near doubling of this city's crime rate! That is because the median does not get pushed and pulled in various directions when there are extraordinarily high or low values in the data set. As we will see, this feature of the median gives it an edge over the mean, the latter of which *is* sensitive to extremely high or extremely low values and does shift accordingly.

The median has the disadvantage of not fully using all available data points. The median offers more information than the mode does, but it still does not account for the entire array of data. This shortfall of the median can be seen by going back to the previous example regarding Minot's property crime rate. The fact that the median did not change when the endpoint of the distribution was noticeably altered demonstrates how the median fails to offer a comprehensive picture of the entire data set. Another disadvantage of the median is that it usually cannot be employed in further statistical computations. There are limited exceptions to this rule, but, generally speaking, the median cannot be plugged into statistical formulas for purposes of performing more-complex analyses.

The Mean

Mean: The arithmetic average of a set of data.

This brings us to the third measure of central tendency that we will cover: the **mean**. The mean is the arithmetic average of a data set. Unlike locating the median, calculating the mean requires using every raw score in a data set. Each individual point exerts a separate and independent effect on the value of the mean. The mean can be calculated only with continuous (interval or ratio) data; it cannot be used to describe categorical variables.

There are two formulas for the computation of the mean, each of which is for use with a particular type of data distribution. The first formula is one with which you are likely familiar from college or high school math classes. The formula is

$$\bar{x} = \frac{\sum x}{N}, \qquad \text{Formula 4(2)}$$

where

$\bar{x}$ (x bar) = the sample mean,

Σ (*sigma*) = a summation sign directing you to sum all numbers or terms to the right of it,

x = values in a given data set, and

N = the sample size.

This formula tells you that to compute the mean, you must first add all the values in the data set together and then divide that sum by the total number of values. Division is required because, all else being equal, larger data sets will produce larger sums, so it is vital to account for sample size when attempting to construct a composite measure such as the mean.

For the example concerning computation of the mean, we can reuse the Colorado violent crime rate data from Table 4.2:

$$\bar{x} = \frac{8.77 + 43.83 + 67.39 + 15.30 + 22.74}{5} = \frac{158.03}{5} = 31.61$$

In 2015, these five cities had a mean violent crime rate of 31.61 per 10,000 residents. Let us try one more example using data from Table 4.4 (homicide rates in 8 states). The mean is calculated as

$$\bar{x} = \frac{7.16 + 6.08 + 5.79 + 1.33 + 1.93 + 5.63 + 4.40 + 1.88}{8} = \frac{34.20}{8} = 4.28$$

Learning Check 4.6

Practice calculating the mean using the property crime rates in Table 4.3.

The second formula for the mean is used for large data sets that are organized in tabular format using both an x column that contains the raw scores and a frequency (f) column that conveys information about how many times each x value occurs in the data set. Table 4.7 shows data from the Bureau of Justice Statistics (BJS) on the number of death-sentenced prisoners received per state in 2013 (Snell, 2014). Note that the f column sums to 36 rather than 50 because in 2013, 14 states did not authorize the death penalty and are thus excluded from the analysis. (This number climbed to 15 that same year when Maryland abolished capital punishment and the governor commuted the sentences of the four people remaining on death row.) Table 4.7 contains the numbers that states reported receiving and the frequency of each number. For instance, 20 states that authorize the death penalty did not admit any new prisoners to death row, while 5 admitted one new inmate each.

To calculate the mean using frequencies, first add a new column to the table. This column—titled *fx*—is the product of the *x* and *f* columns. The results of these calculations for the death row data are located in the right-hand column of Table 4.8. The *fx* column saves time by using multiplication as a shortcut and thereby avoiding cumbersome addition. Using the conventional mean formula would require extensive addition because you would have to sum 36 numbers (i.e., 20 zeroes plus 5 ones plus 4 twos, and so on). This process is unwieldy and impractical, particularly with very large data sets. Instead, merely multiply each value by its frequency and then sum these products to find the total sum of all *x* values. You can see from Table 4.8 that, in 2013, states received 81 new death-sentenced inmates.

Table 4.7 Number of Death-Sentenced Prisoners Received by States, 2013

Number Received (x)	f
0	20
1	5
2	2
3	2
4	3
5	1
9	1
15	1
25	1
	$N = 36$

Once the *fx* column is complete and has been summed, the mean can be calculated using a formula slightly different from the one presented in Formula 4(2). The mean formula for large data sets is

$$\overline{x} = \frac{\sum fx}{N}, \qquad \text{Formula 4(3)}$$

where

f = the frequency associated with each raw score x and

fx = the product of x and f.

Table 4.8 Number of Death-Sentenced Prisoners Received by States, 2013

Number Received (x)	f	fx
0	20	0
1	5	5
2	2	4
3	2	6
4	3	12
5	1	5
9	1	9
15	1	15
25	1	25
	$N = 36$	$\Sigma = 81$

The process of computing this mean can be broken down into three steps: (1) Multiply each *x* by its *f*, (2) sum the resulting *fx* products, and (3) divide by the sample size *N*. Plugging the numbers from Table 4.8 into the formula, it can be seen that the mean is

$$\bar{x} = \frac{81}{36} = 2.25$$

In 2013, the 36 states that authorized use of the death penalty each received a mean of 2.25 new death-sentenced prisoners.

Learning Check 4.7

Anytime you need to compute a mean, you will have to choose between Formulas 4(2) and 4(3). This is a simple enough choice if you just consider that in order to use the formula with an *f* in it, there must be an *f* column in the table. If there is no *f* column, use the formula that does not have an *f*. Refer back to Table 2.9 in Chapter 2. Which formula would be used to calculate the mean? Explain your answer. As practice, use the formula and compute the mean number of prisoners executed per state.

(Continued)

(Continued)

Note that Table 2.9 contains all 50 states, including those states that have abolished capital punishment, whereas Tables 4.7 and 4.8 contain only those 36 states that allowed the death penalty at the time the data were collected. What would happen to the mean calculated based on Table 4.8 if the 14 states that did not allow the death penalty (and thus had zero admissions) were added to the table and to the computation of the mean? Alternatively, what would happen to the mean you calculated on Table 2.9 if the 14 states without capital punishment (which therefore had no executions) were removed from the calculation of the mean? Think about these questions theoretically and make a prediction about whether the mean would increase or decrease. Then make the change to each set of data and redo the means. Were your predictions correct?

Research Example 4.2

How Do Offenders' Criminal Trajectories Impact the Effectiveness of Incarceration?

It is well known that some offenders commit a multitude of crimes over their life and others commit only a few, but the intersection of offense volume (offending rate) and time (the length of a criminal career) has received little attention from criminal justice/criminology researchers. Piquero, Sullivan, and Farrington (2010) used a longitudinal data set of males in South London who demonstrated delinquent behavior early in life and were thereafter tracked by a research team who interviewed them and looked up their official conviction records. The researchers were interested in finding out whether males who committed a lot of crimes in a short amount of time (the short-term, high-rate [STHR] offenders) differed significantly from those who committed crimes at a lower rate over a longer time (the long-term, low-rate [LTLR] offenders) on criminal justice outcomes. The researchers gathered the following descriptive statistics. The numbers not in parentheses are means. The numbers in parentheses are standard deviations, which we will learn about in the next chapter.

You can see from the table that the LTLR offenders differed from the STHR offenders on a number of dimensions. They were, overall, older at the time of their first arrest and had a longer mean career length. They committed many fewer crimes per year and were much less likely to have been sentenced to prison.

Piquero et al.'s (2010) analysis reveals a dilemma about what should be done about these groups of offenders with respect to sentencing; namely, it shows how complicated the question of imprisonment is. The STHR offenders might appear to be the best candidates for incarceration based on their volume of criminal activity, but these offenders' criminal careers are quite short. It makes no sense from a policy and budgetary perspective to imprison people who would not be committing crimes if they were free in society. The STHR offenders also tended to commit property offenses rather than violent ones. The LTLR offenders, by contrast, committed a disproportionate number of violent offenses despite the fact that their overall number of lifetime offenses was lower than that for the STHR group. Again, though, the question of the utility of imprisonment arises: Is it worth incarcerating someone who, though he might still have many years left in his criminal career, will commit very few crimes during that career? The dilemma of sentencing involves the balance between public safety and the need to be very careful in the allotting of scarce correctional resources.

Variable	*Offender Type*	
	LTLR (N = 44)	*STHR (N = 21)*
Overall career length (years)	14.5 (6.50)	10.8 (4.40)
Offenses committed per year	.42 (.31)	1.25 (.72)
Age at first conviction	17.8 (4.70)	13.5 (2.30)
Total convictions	4.9 (2.40)	11.5 (4.10)
Percentage ever incarcerated	9.1%	61.9%
Number of times incarcerated	1.46 (.50)	1.23 (.66)
Years incarcerated	1.23 (.76)	1.05 (.56)

Source: Adapted from Table 1 in Piquero et al. (2010).

Can Good Parenting Practices Reduce the Criminogenic Impact of Youths' Time Spent in Unstructured Activities?

Youths spending time with peers, away from parents and other adults, are at elevated risk for engaging in delinquency either individually or in a group setting. The chances of unstructured, unsupervised activities leading to antisocial acts increases in disadvantaged urban settings, where opportunities for deviance are higher and youths who have never been in trouble before are more likely to encounter delinquent peers. Janssen, Weerman, and Eichelsheim (2017) posited that parents can reduce the chances that their children's unstructured time will result in deviance or delinquency. The researchers hypothesized that strong bonds between parents and children mitigate the criminogenic impacts of time spent in urban environments with peer groups, as does the extent to which parents monitor their children and set boundaries. They used a longitudinal research design, wherein a random sample of adolescents was interviewed twice over a span of two years. Data were collected on how active parents were in their children's lives, the quality of each parent–child relationship, time spent in unstructured activities within disorderly environments, and the number of delinquent acts each child had committed. The table contains the means and the standard deviations for each wave of data collection.

The authors found that all three measures of positive parenting practices significantly reduced children's delinquency. There was no indication that positive parenting buffers children against the deleterious impacts of criminogenic environments; however, good parenting and strong parent–child bonds did negate the effects of increases in the amount of time adolescents spent in these environments. These results suggest that

(Continued)

(Continued)

Variable	Wave 1		Wave 2	
	Mean	SD	*Mean*	SD
Parental monitoring	17.35	2.66	16.36	4.56
Parental limit setting	16.86	2.66	16.61	2.25
Quality of relationship	23.05	3.30	22.76	3.42
Time spent in unstructured activities	0.73	1.93	1.22	2.43
Delinquency	5.50	9.00	4.31	7.15

Source: Adapted from Table 1 in Janssen et al. (2017).

although parenting practices alone do not fully combat the bad influence of unstructured time spent in disorderly neighborhoods, they can offset the effects of an increase in time spent in this manner. Parents have an important role in preventing their children from slipping into deviance and delinquency.

Learning Check 4.8

Explain why the mean cannot be calculated on ordinal data. Look back at the driving frequency variable in Table 4.6 and identify the information that is missing and prevents you from being able to calculate a mean.

The mean is sensitive to extreme values and outliers, which gives it both an advantage and a disadvantage relative to the median. The advantage is that the mean uses the entire data set and accounts for “weird” values that sometimes appear at the high or low ends of the distribution. The median, by contrast, ignores these values. The disadvantage is that the mean’s sensitivity to extreme values makes this measure somewhat unstable; it is vulnerable to the disproportionate impact that a small number of extreme values can exert on the data set.

To illustrate this property of the mean, consider Table 4.9, which contains the 2015 homicide rates for six cities in California. Trace the changes in the mean homicide rates from left to right. Do you notice how the rate increases with the successive introductions of Los Angeles, Soledad, and San Bernardino? Los Angeles pulls the mean up from

2.17 to 3.41, and Soledad tugs it to 5.11. The most dramatic increase occurs when San Bernardino is added: The mean shoots up to 7.65. The successive addition of higher-rate cities caused, in total, more than a threefold increase from the original mean across the three low-crime cities.

Table 4.9 Homicide Rates in California Cities

Urban Area	*Homicides per 100,000*	*Homicides per 100,000*	*Homicides per 100,000*	*Homicides per 100,000*
San Diego	2.64	2.64	2.64	2.64
Redlands	2.81	2.81	2.81	2.81
Santa Monica	1.07	1.07	1.07	1.07
Los Angeles	–	7.12	7.12	7.12
Soledad	–	–	11.92	11.92
San Bernardino	–	–	–	20.33
Sample size	$N = 3$	$N = 4$	$N = 5$	$N = 6$
Mean	$\bar{x} = \frac{6.52}{3} = 2.17$	$\bar{x} = \frac{13.64}{4} = 3.41$	$\bar{x} = \frac{25.56}{5} = 5.11$	$\bar{x} = \frac{45.89}{6} = 7.65$

This demonstrates how the inclusion of extreme values can cause the mean to move in the direction of those values. A score that is noticeably greater than the others in the sample can draw the mean upward, while a value that is markedly lower than the rest can drive the mean downward. There is a good reason why, for example, average income in the United States is reported as a median rather than a mean—a mean would lump extremely poor people who are barely scraping by in with multibillionaires. That would not be accurate at all! The apparent "average" income in the United States would be huge. Finding the point at which 50% of households sit below that particular annual income and 50% above it is more useful and accurate. Because of its sensitivity to extreme values, the mean is most informative when a distribution is normally distributed; the accuracy of the mean as a true measure of central tendency is reduced in distributions that are positively or negatively skewed, and the mean can actually become not merely inaccurate but downright misleading. For instance, a severely economically divided city could have a mean income in the hundreds of thousands of dollars, even if a significant portion of the local population is impoverished.

Another implication of the mean's sensitivity to extreme values is that the mean and the median can be compared to determine the shape of a distribution, as described in the following section.

Table 4.10 Summary of the Measures of Central Tendency Available for Each Level of Measurement

	Mode	*Median*	*Mean*
Nominal	✓		
Ordinal	✓	✓	
Interval	✓	✓	✓
Ratio	✓	✓	✓

Using the Mean and the Median to Determine Distribution Shape

Given that the median is invulnerable to extreme values but the mean is not, the best strategy is to report both of these measures when describing data distributions. The mean and the median can, in fact, be compared to form a judgment about whether the data are normally distributed, positively skewed, or negatively skewed. In normal distributions, the mean and median will be approximately equal. Positively skewed distributions will have means markedly greater than their medians. This is because extremely high values in positively skewed distributions pull the mean up but do not affect the location of the median. Negatively skewed distributions, on the other hand, will have medians that are noticeably larger than their means because extremely low numbers tug the mean downward but do not alter the median's value. Figure 4.6 illustrates this conceptually.

To give a full picture of a data distribution, then, it is best to make a habit of reporting both the mean and the median.

The mean—unlike the mode or the median—forms the basis for further computations; in fact, the mean is an analytical staple of many inferential hypothesis tests. The reason that the mean can be used in this manner is that the mean is the **midpoint of the magnitudes**. This point is important and merits its own section.

Midpoint of the magnitudes: The property of the mean that causes all deviation scores based on the mean to sum to zero.

Figure 4.6 The Mean and Median as Indicators of Distribution Shape

Deviation Scores and the Mean as the Midpoint of the Magnitudes

Deviation score: The distance between the mean of a data set and any given raw score in that set.

The mean possesses a vital property that enables its use in complex statistical formulas. To understand this, we must first discuss **deviation scores**. A deviation score is a given data point's distance from its group mean. The formula for a deviation score is based on simple subtraction:

$$d_i = x_i - \bar{x},$$ *Formula 4(4)*

where

d_i = the deviation score for a given data point x_i, and
x_i = a given data point,
$\bar{x}$ = the sample mean.

Suppose, for instance, that a group's mean is $\bar{x} = 24$. If a certain raw score x_i is 22, then $d_{x=22} = 22 - 24 = -2$. A raw score of 25, by contrast, would have a deviation score of $d_{x=25} = 25 - 24 = 1$.

A deviation score conveys two pieces of information. The first is the absolute value of the score or, in other words, how far from the mean a particular raw score is. Data points that are exactly equal to the mean will have deviation scores of 0; therefore, deviation scores with larger absolute values are farther away from the mean, while deviation scores with smaller absolute values are closer to it.

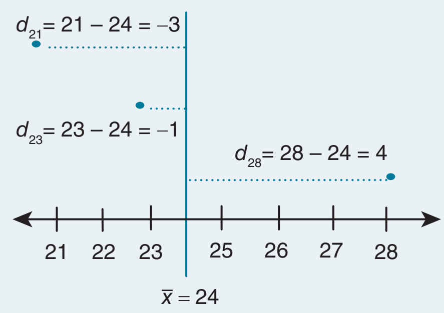
The second piece of information that a deviation score conveys is whether the raw score associated with that deviation score is greater than or less than the mean. A deviation score's sign (positive or negative) communicates this information. Positive deviation scores represent raw scores that are greater than the mean and negative deviation scores signify raw numbers that are less than the mean. You can thus discern two characteristics of the raw score x_i that a given deviation score d_i represents: (1) the distance between x_i and $\bar{x}$ and (2) whether x_i is above $\bar{x}$ or below it. Notice that you would not even need to know the actual value of x_i or $\bar{x}$ in order to effectively interpret a deviation score. Deviation scores convey information about the position of a given data point with respect to its group mean; that is, deviation scores offer information about raw scores' *relative*, rather than absolute, positions within their group. Figure 4.7 illustrates this.

What lends the mean its title as the midpoint of the magnitudes is the fact that deviation scores computed using the mean (as opposed to the mode or median) always sum to zero (or within rounding error of it). The mean is the value in the data set at which all values below it balance out with all values above it. For an example of this, try summing the deviation scores in Figure 4.7. What is the result?

To demonstrate this concept more concretely, the raw homicide counts that were used as rates in Table 4.9 are listed in Table 4.11. These cities had a mean of 87.25 homicides in 2015. The *d* column contains the deviation score for each homicide count.

Illustrative of the mean as the fulcrum of the data set, the positive and negative deviation scores ultimately cancel each other out, as can be seen by the sum of zero at the bottom

Figure 4.7 Deviation Scores in a Set of Data With a Mean of 24



of the deviation-score column. This represents the mean's property of being the midpoint of the magnitudes—it is the value that perfectly balances all of the raw scores. This characteristic is what makes the mean a central component in more-complex statistical analyses. You will see in later chapters that the mean features prominently in many calculations.

Table 4.11 Homicides in California Cities

City	*Homicides*	d
San Diego	37	37 – 63.50 = –26.50
Redlands	2	2 – 63.50 = –61.50
Santa Monica	13	13 – 63.50 = –50.50
Los Angeles	282	282 – 63.50 = 218.50
Soledad	3	3 – 63.50 = –60.50
San Bernardino	44	44 – 63.50 = –19.50
	$\bar{x} = \frac{381}{6} = 63.50$	$\Sigma = 0.00$

Learning Check 4.9

To test your comprehension of the concept of the mean as the midpoint of the magnitudes, go back to Table 4.3. You calculated the mean property-crime rate in Learning Check 4.6. Use that mean to compute each city's deviation score, and then sum the scores.

SPSS

Criminal justice and criminology researchers generally work with large data sets, so computing measures of central tendency by hand is not feasible; luckily, it is not necessary, either, because statistical programs such as SPSS can be used instead. There are two different ways to obtain central tendency output. Under the *Analyze* → *Descriptive Statistics* menu, SPSS offers the options *Descriptives* and *Frequencies*. Both of these functions will produce central tendency analyses, but the *Frequencies* option offers a broader array of descriptive statistics and even some charts and graphs. For this reason, we will use *Frequencies* rather than *Descriptives*. Once you have opened the *Frequencies* box, click on the *Statistics* button to open a menu of options for measures of central tendency. Select *Mean*, *Median*, and *Mode*, as shown in Figure 4.8. Then click *OK*, and the output displayed in Figure 4.9 will appear. The variable used in this example comes from Table 4.7, measuring the number of prisoners received under sentence of death in 2015.

You can see from Figure 4.9 that the mean is identical to the result we arrived at by hand earlier. The mode is zero, which you can verify by looking at Table 4.7. The median is zero, meaning half the states did not receive any new death-row inmates and half received one or more. We can also compare the mean and the median to determine the shape of the distribution. With a mean of 2.25 and a median of zero, do you think that this distribution is normally distributed, positively skewed, or negatively skewed? If you said positively skewed, you are correct!

Figure 4.8 Running Measures of Central Tendency in SPSS

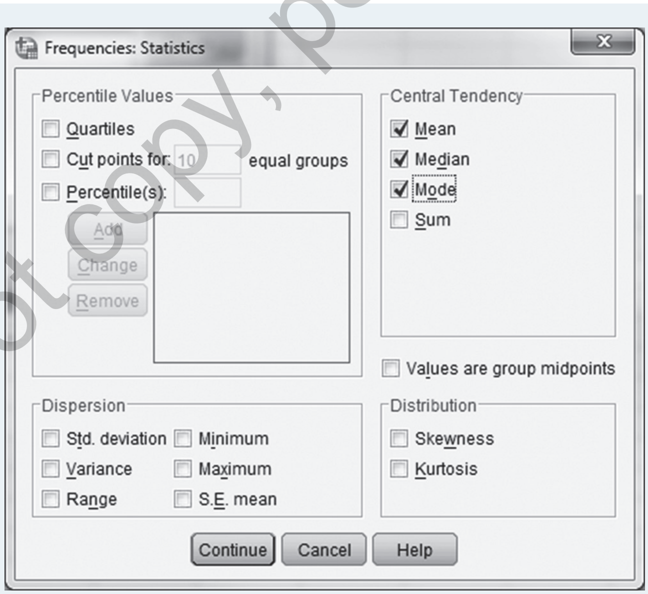


Figure 4.9 SPSS Output

Frequencies

Statistics

Prisoners received 2013

N	Valid	36
	Missing	0
Mean		2.2500
Median		.0000
Mode		.00

For another example, we will use 2015 homicide rates (per 100,000) in all California cities. Figure 4.10 shows the mean, the median, and the mode for this variable.

Follow along with this example of using SPSS to obtain measures of central tendency by downloading the file *California Homicides for Chapter 4.sav* at **edge.sagepub.com/gau3e.**

Figure 4.10 Homicide Rates in California Cities

Frequencies

Statistics

California homicide rates, 2015

N	Valid	460
	Missing	0
Mean		6.5782
Median		1.0839
Mode		.00

The mode is zero, but because homicide rates vary so widely across this large sample of cities (N = 460), the mode is not a useful or informative measure. More interesting are the mean and the median. The mean (6.58) is much larger than the median (1.08), indicating significant positive skew. This can be verified by using SPSS to produce a histogram of the data using the Chart Builder; refer back to Chapter 3 if you need a refresher on the use of the Chart Builder. Figure 4.11 shows the histogram.

Figure 4.11 Histogram of Homicide Rates in California Cities

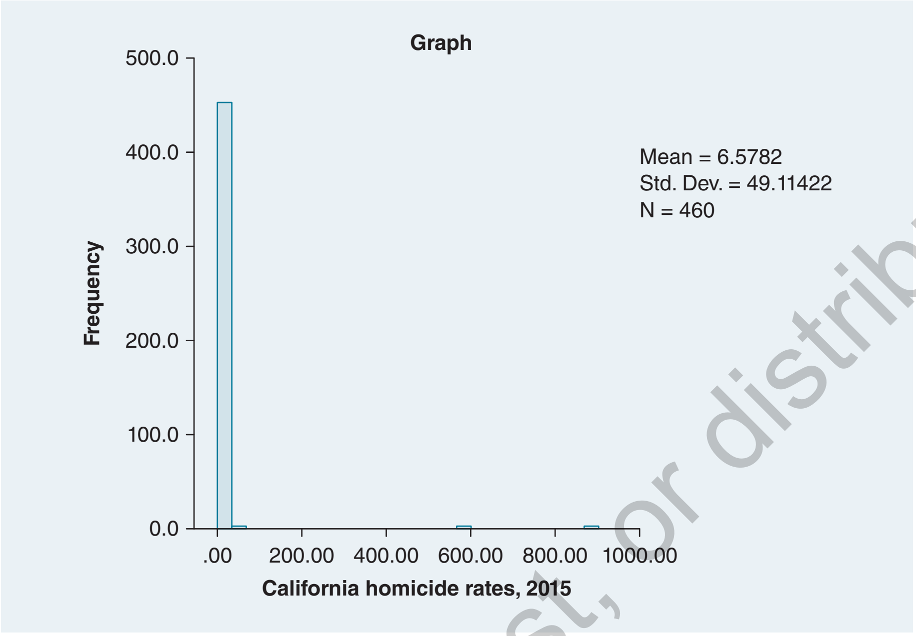


Figure 4.11 confirms the severe skew in this variable: The data cluster toward the lower end so much that the values out in the tail are barely visible in the histogram.

Learning Check 4.10

Flip back a few pages to Figure 4.1. Recall that this is an example of a normal distribution. Based on the distribution's shape, how close or far apart do you think the mean and the median are? In other words, do you think they are close together, or do you predict that one is a lot bigger (or smaller) than the other? Explain your answer.

There is a GIGO alert relevant here. It is your responsibility to ensure that you use the correct measure(s) of central tendency given the level of measurement of the variable with which you are working. The SPSS program will not produce an error message if you make a mistake by, for instance, telling it to give you the mean of a nominal or ordinal variable. You will get a mean, just the same as you get when you correctly ask for the mean of a continuous variable. To illustrate this, Figure 4.12 contains output from the National Crime Victimization Survey (NCVS) showing respondents' marital status. Although this is a nominal variable—making the median and the mean inappropriate—SPSS went ahead with the calculations anyway and produced results. Of course, the mean and the median of a variable measuring whether someone is married, separated, divorced, and so on are nonsense, but SPSS does not know that. This statistical program is not a substitute for knowing which techniques are appropriate for which data types.

Figure 4.12 Respondent Marital Status

Statistics

Marital status

N	Valid	84229
	Missing	594
Mean		2.07
Median		1.00
Mode		1

CHAPTER SUMMARY

This chapter introduced you to three measures of central tendency: mode, median, and mean. These statistics offer summary information about the middle or average score in a data set. The mode is the most frequently occurring category or value in a data set. The mode can be used with variables of any measurement type (nominal, ordinal, interval, or ratio) and is the only measure that can be used with nominal variables. Its main weakness is in its simplicity and superficiality—it is generally not all that useful.

The median is a better measure than the mode for data measured at the ordinal or continuous level. The median is the value that splits a data set exactly in half. Since it is a positional measure, the median's value is not affected by the presence of extreme values; this makes the median a better reflection of the center of a distribution the mean is when a distribution is highly skewed. The median, though, does not take into account all data points in a distribution, which makes it less informative than the mean.

The mean is the arithmetic average of the data and is used with continuous variables only. The mean accounts for all values in a data set, which is good because no data are omitted; the flipside, however, is that the mean is susceptible to being pushed and pulled by extreme values.

It is good to report both the mean and the median because they can be compared to determine the shape of a distribution. In a normal distribution, these two statistics will be approximately equal; in a positively skewed distribution, the mean will be markedly greater than the median; and in a negatively skewed distribution, the mean will be noticeably smaller than the median. Reporting both of them provides your audience with much more information than they would have if you just reported one or the other.

The mode, the median, and the mean can all be obtained in SPSS using the *Analyze* → *Descriptive Statistics* → *Frequencies* sequence. As always, GIGO!

When you order SPSS to produce a measure of central tendency, it is your responsibility to ensure that the measure you choose is appropriate to the variable's level of measurement. If you err, SPSS will probably not alert you to the mistake—you will get output that looks fine but is actually garbage. Be careful!

THINKING CRITICALLY

1. According to the Law Enforcement Management and Administrative Statistics (LEMAS) survey, the mean number of officers per police agency (of all types) is 163.92. Do you trust that this mean is an accurate representation of the middle of the distribution of police agency size? Why or why not? If not, what additional information would you need in order to gain an accurate understanding of this distribution's shape and central tendency?
2. The COJ reports that the modal number of inmates per jail is 1. This value occurs more frequently than any other population value (51 times among 2,371 jails). Use this example to discuss the limitations and drawbacks of using the mode to describe the central tendency of a continuous variable. Then identify the measure(s) you would use instead of the mode, and explain why.

REVIEW PROBLEMS

1. A survey item asks respondents, "How many times have you shoplifted?" and allows them to fill in the appropriate number.
   a. What level of measurement is this variable?
   b. What measure or measures of central tendency can be computed on this variable?
2. A survey item asks respondents, "How many times have you shoplifted?" and gives them the answer options: *0, 1–3, 4–6, 7 or more*.
   a. What level of measurement is this variable?
   b. What measure or measures of central tendency can be computed on this variable?
3. A survey item asks respondents, "Have you ever shoplifted?" and tells them to circle *yes* or *no*.
   a. What level of measurement is this variable?
   b. What measure or measures of central tendency can be computed on this variable?
4. Explain what an extreme value is. Include in your answer (1) the effect extreme values have on the median, if any, and (2) the effect extreme values have on the mean, if any.
5. Explain why the mean is the midpoint of the magnitudes. Include in your answer (1) what deviation scores are and how they are calculated and (2) what deviation scores always sum to.
6. In a negatively skewed distribution . . .
   a. the mean is less than the median.
   b. the mean is greater than the median.
   c. the mean and the median are approximately equal.
7. In a normal distribution . . .
   a. the mean is less than the median.
   b. the mean is greater than the median.
   c. the mean and the median are approximately equal.
8. In a positively skewed distribution . . .
   a. the mean is less than the median.
   b. the mean is greater than the median.
   c. the mean and the median are approximately equal.

9. In a positively skewed distribution, the tail extends toward _____ of the number line.

   a. the positive side
   b. both sides
   c. the negative side

10. In a negatively skewed distribution, the tail extends toward _____ of the number line.

   a. the positive side
   b. both sides
   c. the negative side

11. The following table contains 2015 UCR data on the relationship between murder victims and their killers, among those crimes for which the relationship status is known.

Victim–Offender Relationships Among Murder Victims

Offender's Relationship to Victim	*f*
Family Member	1,099
Intimate Partner	1,270
Friend	365
Acquaintance	2,801
Other Nonstranger	113
Stranger	1,375
	N = 7,023

   a. Identify this variable's level of measurement and, based on that, state the appropriate measure or measures of central tendency.
   b. Determine or calculate the measure or measures of central tendency that you identified in part (a).

12. The frequency distribution in the following table shows rates of violent victimization, per victim racial group, in 2014 according to the NCVS (Truman & Langton, 2015). Use this table to do the tasks following the table.

Violent Victimization Rate per 1,000 Persons

Race	*Victimization Rate*
Black/African American	22.5
White	20.3
Hispanic/Latino	16.2
Other	23.0
N = 4	

   a. Identify the median victimization rate using all three steps.
   b. Compute the mean victimization rate across all racial groups.

13. Morgan, Morgan, and Boba (2010) report state and local government expenditures, by state, for police protection in 2007. The data in the following table contain a random sample of states and the dollars spent per capita in each state for police services.

Dollars Spent per Capita on Police Protection, 2007

State	*Dollars Spent per Capita*
Illinois	317
Arkansas	170
Alabama	211
Ohio	258
Washington State	219
Florida	345
Maine	176
Texas	220
N = 8	

   a. Identify the median dollar amount using all three steps.
   b. Calculate the mean dollar amount.

14. The following frequency distribution shows LEMAS data on the number of American Indian officers employed by state police agencies.

American Indian Officers in State Police Agencies

Number of Officers	f	Number of Officers	f
0	8	13	1
1	4	14	1
2	3	15	1
3	2	19	1
4	3	20	1
6	2	27	1
7	4	40	1
9	3	44	1
10	1	45	1
11	1	83	1
12	1		$N = 42$

a. Identify the modal number of American Indian officers in this sample of agencies.

b. Compute the mean number of American Indian officers in this sample.

c. The median number of American Indian officers is 6.00. Based on this median and the mean you calculated, would you say that this distribution is normally distributed, positively skewed, or negatively skewed? Explain your answer.

15. The following frequency distribution shows a variable from the PPCS measuring, among female respondents who had been stopped by police while walking or riding a bike, the number of minutes those stops lasted.

Number of Minutes Stop Lasted

Minutes	f	Number	f
1	15	20	1
2	4	24	1
3	2	30	3
4	2	45	2
5	15	60	2
7	1	120	2
8	1	180	1
10	11		$N = 70$
15	7		

a. Compute the mean number of minutes per stop.

b. The median number of minutes was 5.00. Based on this median and the mean you calculated earlier, would you say that this distribution is normally distributed, positively skewed, or negatively skewed? Explain your answer.

16. The following table shows UCR data on the number of aggravated assaults that occurred in five Wyoming cities and towns in 2015.

Aggravated Assaults in Wyoming Cities and Towns

City or Town	Agg. Assaults
Pine Bluffs	6
Jackson	14
Laramie	32
Afton	1
Cheyenne	95
$N = 5$	

a. Identify the median number of assaults in this sample.

b. Calculate the mean number of assaults.

c. Calculate each city's deviation score, and sum the scores.

17. The following table displays the number of juveniles arrested for arson in select states in 2015, according to the UCR.

Juveniles Arrested for Arson

State	Arrests
Alaska	31
Arkansas	26
Connecticut	31
Kansas	15
Montana	13
South Carolina	27
Vermont	4
$N = 7$	

a. Identify the median number of arson arrests per state.

b. Calculate the mean number of arrests in this sample.

c. Calculate each state's deviation score and sum the scores.

18. The data set *NCVS for Chapter 4.sav* **(edge .sagepub.com/gau3e)** contains the ages of respondents to the 2015 NCVS. Run an SPSS analysis to determine the mode, the median, and the mean of this variable. Summarize the results in words.

> Helpful Hint: When running measures of central tendency on large data sets in SPSS, deselect the "Display frequency tables" option in the "Frequencies" dialog box. This will not alter the analyses you are running but will make your output cleaner and simpler to examine.

19. The data set *NCVS for Chapter 4.sav* **(edge .sagepub.com/gau3e)** also contains portions of the Identity Theft Supplement survey. The variable "purchases" asked respondents how many times they had made online purchases in the past year. Run an SPSS analysis to determine the mode, the median, and the mean of this variable. Summarize the results in words.

20. The data file *Staff Ratio for Chapter 4.sav* **(edge .sagepub.com/gau3e)** contains a variable from the 2006 COJ showing the ratio of inmates to security staff per institution. Run an SPSS analysis to determine the mode, the median, and the mean of this variable. Summarize the results in words.

KEY TERMS

Measures of central tendency 76
Normal distribution 76
Positive skew 77
Negative skew 77
Mode 78
Median 83
Mean 88
Midpoint of the magnitudes 96
Deviation score 97

GLOSSARY OF SYMBOLS AND ABBREVIATIONS INTRODUCED IN THIS CHAPTER

MP	Median position
Md	Median
$\overline{X}$	Mean
d_i	Deviation score for a raw score x_i