OBJECTIVES

1. Display data using bar graphs, broken-line graphs, and pie graphs.
2. Display quantitative data using histograms, frequency polygons, and cumulative frequency polygons.
3. Compute the mean, median, and mode of a data set to measure central tendency.
4. Compute the range, variance, and standard deviation of a data set to measure variation.
5. Calculate the probability distribution of the number of successes in a sequence of Bernoulli trials.
6. Find the mean and standard deviation of a binomial distribution.
7. Calculate probabilities associated with normal distributions.
8. Approximate a binomial distribution with an appropriately chosen normal distribution.

CHAPTER PROBLEM

Most colleges and universities require applicants to submit their scores on at least one of two very popular entrance exams, the ACT (American College Testing Program) or the SAT (Scholastic Assessment Test). In the year that two students, Jack and Jill, took one of these exams, the national mean and standard deviation for the ACT were 20.8 and 4.5, respectively, and the national mean and standard deviation for the SAT were 1173 and 100, respectively. Jack scored 26 on the ACT and Jill scored 1260 on the SAT. What percentage of ACT exam participants scored less than Jack? What percentage of SAT exam participants scored less than Jill on the SAT? Whose percentage is higher, Jack’s or Jill’s?
8-1 Graphing Data
8-2 Measures of Central Tendency
8-3 Measures of Dispersion
8-4 Bernoulli Trials and Binomial Distributions
8-5 Normal Distributions
Chapter 8 Review
Review Exercise
Group Activity 1: Analysis of Data on Student Lifestyle
Group Activity 2: Survival Rates for a Heart Transplant

INTRODUCTION

In this chapter we study various techniques for analyzing and displaying data. We use bar graphs, broken-line graphs, and pie graphs to present visual interpretations or comparisons of data. We use measures of central tendency (the mean, median, and mode) and measures of dispersion (the range, variance, and standard deviation) to describe and compare data sets.

Data collected from different sources, IQ scores and measurements of manufactured parts, for example, often exhibit surprising similarity. We might express such similarity by saying that both data sets exhibit characteristics of a normal distribution. In this chapter we develop theoretical probability distributions—the binomial distributions and the normal distributions—that can be used as models of empirical data.

Section 8-1

Graphing Data

➤ Bar Graphs, Broken-Line Graphs, and Pie Graphs
➤ Frequency Distributions
➤ Comments on Statistics
➤ Histograms
➤ Frequency Polygons and Cumulative Frequency Polygons

Television, newspapers, magazines, books, and reports make substantial use of graphics to visually communicate complicated sets of data to the viewer. In this section we look at bar graphs, broken-line graphs, and pie graphs and the techniques for producing them. It is important to remember that graphs are
visual aids and should be prepared with care. The object is to provide the viewer with the maximum amount of information while minimizing the time and effort required to “read” the information from the graph.

➤ **Bar Graphs, Broken-Line Graphs, and Pie Graphs**

**Bar graphs** are widely used because they are easy to construct and easy to read. They are effective in presenting visual interpretations or comparisons of data. Consider Tables 1 and 2. Bar graphs are well suited to describe these two data sets. Vertical bars are usually used for time series—that is, data that changes over time, as in Table 1. The labels on the horizontal axis are then units of time (hours, days, years, and so on, whichever is appropriate), as shown in Figure 1. Horizontal bars are generally used for data that changes by category, as in Table 2, because of the ease of labeling categories on the vertical axis of the bar graph (see Fig. 2). To increase clarity, a space is left between the bars. Bar graphs for the data in Tables 1 and 2 are illustrated in Figures 1 and 2.

Two additional variations on bar graphs, the double bar graph and the divided bar graph, are illustrated in Figures 3 and 4, respectively.

<table>
<thead>
<tr>
<th>Year</th>
<th>Debt (billions $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>284.1</td>
</tr>
<tr>
<td>1970</td>
<td>370.1</td>
</tr>
<tr>
<td>1980</td>
<td>907.7</td>
</tr>
<tr>
<td>1990</td>
<td>3,233.3</td>
</tr>
<tr>
<td>2000</td>
<td>5,674.2</td>
</tr>
</tbody>
</table>

**TABLE 1 U.S. Public Debt**

**TABLE 2 Traffic at Busiest U.S. Airports, 2001**

<table>
<thead>
<tr>
<th>Airport</th>
<th>Arrivals and Departures (million passengers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>76</td>
</tr>
<tr>
<td>Chicago (O’Hare)</td>
<td>67</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>61</td>
</tr>
<tr>
<td>Dallas/Ft. Worth</td>
<td>55</td>
</tr>
<tr>
<td>Denver</td>
<td>36</td>
</tr>
</tbody>
</table>

**FIGURE 1 Vertical bar graph**

**FIGURE 2 Horizontal bar graph**
(A) Using Figure 3, estimate the mean annual income of a male with some college, and of a female who holds a bachelor’s degree. Within which educational category is there the greatest difference between male and female income? The least difference?

(B) Using Figure 4, estimate the population of São Paulo in the years 2000 and 2015. Which of the six largest cities is projected to have the greatest increase in population from 2000 to 2015? The least increase? Which city would you conjecture to be the largest in the world in 2035? Explain.

A **broken-line graph** can be obtained from a vertical bar graph by joining the midpoints of the tops of consecutive bars with straight lines. For example, using Figure 1 we obtain the broken-line graph in Figure 5.
Broken-line graphs are particularly useful when we want to emphasize the change in one or more variables relative to time. Figures 6 and 7 illustrate two additional variations of broken-line graphs.

(A) Using Figure 6, estimate the revenue and costs in 2000. In which years is a profit realized? In which year is the greatest loss experienced?

(B) Using Figure 7, estimate the U.S. consumption of each of the five sources of energy in 2010. Estimate the percentage of total consumption that will come from nuclear energy in the year 2010.

A pie graph is generally used to show how a whole is divided among several categories. The amount in each category is expressed as a percentage, and then a circle is divided into segments (pieces of pie) proportional to the percentages of each category. The central angle of a segment is the percentage of 360° corresponding to the percentage of that category (see Fig. 8). In constructing pie graphs, we use relatively few categories, arrange the segments in ascending or descending order of size around the circle, and label each part.

Bar graphs, broken-line graphs, and pie graphs are easily constructed using a spreadsheet. After the data is entered (see Fig. 9 for the data corresponding
to Fig. 8A) and the type of display (bar, broken-line, pie) is chosen, the graph is drawn automatically. Various options for axes, gridlines, patterns, and text are available to improve the clarity of the visual display.

### FIGURE 9

#### Frequency Distributions

Observations that are measured on a numerical scale are referred to as **quantitative data**. Weights, ages, bond yields, the length of a part in a manufacturing process, test scores, and so on, are all examples of quantitative data. Out of the total population of entering freshmen at a large university, a random sample of 100 students is selected and their entrance examination scores are recorded (see Table 3).

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Army</td>
<td>Navy</td>
<td>Air Force</td>
<td>Marine Corps</td>
<td>Coast Guard</td>
</tr>
<tr>
<td>2</td>
<td>485536</td>
<td>384576</td>
<td>369721</td>
<td>173385</td>
<td>37166</td>
</tr>
</tbody>
</table>

The mass of raw data in Table 3 certainly does not elicit much interest or exhibit much useful information. The data must be organized in some way so that it is comprehensible. This can be done by constructing a frequency table.

We generally choose five to twenty **class intervals** of equal length to cover the data range—the more data, the greater the number of intervals—and tally the data relative to these intervals. The **data range** in Table 1 is $787 - 340 = 447$ (found by subtracting the smallest value in the data from the largest). If we choose ten intervals, each of length 50, we will be able to cover all the scores. Table 4 shows the result of this tally.

### TABLE 3 Entrance Examination Scores of 100 Entering Freshmen

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>299.5–349.5</td>
<td></td>
<td>1</td>
<td>.01</td>
</tr>
<tr>
<td>349.5–399.5</td>
<td></td>
<td>2</td>
<td>.02</td>
</tr>
<tr>
<td>399.5–449.5</td>
<td></td>
<td>5</td>
<td>.05</td>
</tr>
<tr>
<td>449.5–499.5</td>
<td></td>
<td>10</td>
<td>.10</td>
</tr>
<tr>
<td>499.5–549.5</td>
<td></td>
<td>21</td>
<td>.21</td>
</tr>
<tr>
<td>549.5–599.5</td>
<td></td>
<td>20</td>
<td>.20</td>
</tr>
<tr>
<td>599.5–649.5</td>
<td></td>
<td>19</td>
<td>.19</td>
</tr>
<tr>
<td>649.5–699.5</td>
<td></td>
<td>11</td>
<td>.11</td>
</tr>
<tr>
<td>699.5–749.5</td>
<td></td>
<td>7</td>
<td>.07</td>
</tr>
<tr>
<td>749.5–799.5</td>
<td></td>
<td>4</td>
<td>.04</td>
</tr>
</tbody>
</table>

### TABLE 4 Frequency Table

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>299.5–349.5</td>
<td></td>
<td>1</td>
<td>.01</td>
</tr>
<tr>
<td>349.5–399.5</td>
<td></td>
<td>2</td>
<td>.02</td>
</tr>
<tr>
<td>399.5–449.5</td>
<td></td>
<td>5</td>
<td>.05</td>
</tr>
<tr>
<td>449.5–499.5</td>
<td></td>
<td>10</td>
<td>.10</td>
</tr>
<tr>
<td>499.5–549.5</td>
<td></td>
<td>21</td>
<td>.21</td>
</tr>
<tr>
<td>549.5–599.5</td>
<td></td>
<td>20</td>
<td>.20</td>
</tr>
<tr>
<td>599.5–649.5</td>
<td></td>
<td>19</td>
<td>.19</td>
</tr>
<tr>
<td>649.5–699.5</td>
<td></td>
<td>11</td>
<td>.11</td>
</tr>
<tr>
<td>699.5–749.5</td>
<td></td>
<td>7</td>
<td>.07</td>
</tr>
<tr>
<td>749.5–799.5</td>
<td></td>
<td>4</td>
<td>.04</td>
</tr>
</tbody>
</table>

The data range in Table 1 is $787 - 340 = 447$. If we choose ten intervals, each of length 50, we will be able to cover all the scores. Table 4 shows the result of this tally.
At first it might seem appropriate to start at 300 and form the class inter-
vals: 300–350, 350–400, 400–450, and so on. But if we do this, where will we
place 350 or 400? We could, of course, adopt a convention of placing a score
falling on an upper boundary of a class in the next higher class (and some
people do exactly this); however, to avoid confusion, we will always use one deci-
mal place more for class boundaries than appears in the raw data. Thus, in this
case, we chose the class intervals 299.5–349.5, 349.5–399.5, and so on, so that
each score could be assigned to one and only one class interval.

The number of measurements that fall within a given class interval is called
the class frequency, and the set of all such frequencies associated with their
corresponding classes is called a frequency distribution. Thus, Table 4 repre-
sents a frequency distribution of the set of raw scores in Table 3. If we divide
each frequency by the total number of items in the original data set (in our
case 100), we obtain the relative frequency of the data falling in each class in-
terval—that is, the percentage of the whole that falls in each class interval (see
the last column in Table 4).

The relative frequencies also can be interpreted as probabilities associated
with the experiment, “A score is drawn at random out of the 100 in the
sample.” An appropriate sample space for this experiment would be the set of
simple outcomes

\[ e_1 = \text{a score falls in the first class interval} \]
\[ e_2 = \text{a score falls in the second class interval} \]
\[ \vdots \]
\[ e_{10} = \text{a score falls in the tenth class interval} \]

The set of relative frequencies is then referred to as the probability distribu-
tion for the sample space.

**Example 1**

**Determining Probabilities from a Frequency Table** Referring to
the probability distribution just described and Table 4, determine the probabil-
ity that

(A) A randomly drawn score is between 499.5 and 549.5.
(B) A randomly drawn score is between 449.5 and 649.5.

**Solution**

(A) Since the relative frequency associated with the class interval 499.5–549.5
is .21, the probability that a randomly drawn score (from the sample of
100) falls in this interval is .21.

(B) Since a score falling in the interval 449.5–649.5 is a compound event, we
simply add the probabilities for the simple events whose union is this
compound event. Thus, we add the probabilities corresponding to each
class interval from 449.5 to 649.5 to obtain

\[ .10 + .21 + .20 + .19 = .70 \]

**Matched Problem 1** Repeat Example 1 for the following intervals:
(A) 649.5–699.5
(B) 299.5–499.5
Comments on Statistics

Now, of course, what we are really interested in is whether the probability distribution for the sample of 100 entrance examination scores has anything to do with the total population of entering freshmen in the university in question. This is a problem for the important branch of mathematics called statistics, which deals with the process of making inferences about a total population based on random samples drawn from the population. Figure 10 schematically illustrates the inferential statistical process. We will not go too far into inferential statistics in this book, since the subject is studied in detail in any course in statistics, but our work in probability provides a good foundation for this study.

Intuitively, in the entrance examination example, we would expect that the larger the sample size, the more closely the probability distribution for the sample will approximate that for the total population. That is about all that we can say at the moment.

Histograms

A histogram is a special kind of vertical bar graph. In fact, if you rotate Table 4 on page 497 counterclockwise 90°, the tally marks in the table take on the appearance of a bar graph. Histograms have no space between the bars, class boundaries are located on the horizontal axis, and frequencies are associated with the vertical axis. Figure 11 is a histogram for the frequency distribution in Table 4. Note that we have included both frequencies and relative frequencies on the vertical scale. You can include either one or the other, or both, depending on what needs to be emphasized. The histogram is the most common graphical representation of frequency distributions.

(A) Draw a histogram for the following data set using a class interval width of 0.5 starting at 0.45.

1.5 1.3 1.4 1.5 2.5 1.5 1.4 1.6
1.4 1.5 1.5 1.4 0.9 2.1 1.4 1.5
(B) Draw a histogram for the same data set using the same class interval width but starting at 0.75.

(C) The histograms of parts (A) and (B) represent the same set of data. How do they differ? Which of the two gives a better description of the data set? Explain.

**EXAMPLE 2**

**Constructing Histograms with a Graphing Utility**

Twenty vehicles were chosen at random upon arrival at a vehicle emissions inspection station, and the time elapsed (in minutes) from arrival to completion of the emissions test was recorded for each of the vehicles:

<table>
<thead>
<tr>
<th>5</th>
<th>12</th>
<th>11</th>
<th>4</th>
<th>7</th>
<th>20</th>
<th>14</th>
<th>8</th>
<th>12</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>15</td>
<td>14</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>20</td>
<td>26</td>
<td>17</td>
</tr>
</tbody>
</table>

(A) Use a graphing utility to draw a histogram of the data, choosing the five class intervals 2.5–7.5, 7.5–12.5, and so on.

(B) What is the probability that for a vehicle chosen at random from the sample, the time required at the inspection station is less than 12.5 minutes? That it exceeds 22.5 minutes?

**Solution**

(A) Various kinds of statistical plots can be drawn by most graphing utilities. To draw a histogram we enter the data as a list, specify a histogram from among the various statistical plotting options, set the window variables, and graph. Figure 12 shows the data entered as a list, the settings of the window variables, and the resulting histogram for a particular graphing calculator. For details consult your manual.

![Figure 12](image)

(B) From the histogram in Figure 12 we see that the first class has frequency 3 and the second has frequency 8. The upper boundary of the second class is 12.5, and the total number of data items is 20. Therefore, the probability that the time required is less than 12.5 minutes is

$$\frac{3 + 8}{20} = \frac{11}{20} = .55$$

Similarly, since the frequency of the last class is 1, the probability that the time required exceeds 22.5 minutes is

$$\frac{1}{20} = .05$$

**Matched Problem 2**

The weights (in pounds) were recorded for 20 kindergarten children chosen at random:

<table>
<thead>
<tr>
<th>51</th>
<th>46</th>
<th>37</th>
<th>39</th>
<th>48</th>
<th>42</th>
<th>41</th>
<th>44</th>
<th>57</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>44</td>
<td>41</td>
<td>50</td>
<td>45</td>
<td>46</td>
<td>34</td>
<td>39</td>
<td>42</td>
<td>44</td>
</tr>
</tbody>
</table>
(A) Use a graphing utility to draw a histogram of the data, choosing the five class intervals 32.5–37.5, 37.5–42.5, and so on.

(B) What is the probability that a kindergarten child chosen at random from the sample weighs less than 42.5 pounds? More than 42.5 pounds?

**Frequency Polygons and Cumulative Frequency Polygons**

A frequency polygon is a broken-line graph where successive midpoints of the tops of the bars in a histogram are joined by straight lines. To draw a frequency polygon for a frequency distribution, you do not need to draw a histogram first; you can just locate the midpoints and join them with straight lines. Figure 13 is a frequency polygon for the frequency distribution in Table 4. If the amount of data becomes very large and we substantially increase the number of classes, the frequency polygon will take on the appearance of a smooth curve called a frequency curve.

![Frequency polygon](image)

If we are interested in how many or what percentage of a total sample lies above or below a particular measurement, a cumulative frequency table and polygon are useful. Using the frequency distribution in Table 4, we accumulate the frequencies by starting with the first class and adding frequencies as we move down the column. The results are shown in Table 5. (How is the last column formed?)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Cumulative Frequency</th>
<th>Relative Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>299.5–349.5</td>
<td>1</td>
<td>1</td>
<td>.01</td>
</tr>
<tr>
<td>349.5–399.5</td>
<td>2</td>
<td>3</td>
<td>.03</td>
</tr>
<tr>
<td>399.5–449.5</td>
<td>5</td>
<td>8</td>
<td>.08</td>
</tr>
<tr>
<td>449.5–499.5</td>
<td>10</td>
<td>18</td>
<td>.18</td>
</tr>
<tr>
<td>499.5–549.5</td>
<td>21</td>
<td>39</td>
<td>.39</td>
</tr>
<tr>
<td>549.5–599.5</td>
<td>20</td>
<td>59</td>
<td>.59</td>
</tr>
<tr>
<td>599.5–649.5</td>
<td>19</td>
<td>78</td>
<td>.78</td>
</tr>
<tr>
<td>649.5–699.5</td>
<td>11</td>
<td>89</td>
<td>.89</td>
</tr>
<tr>
<td>699.5–749.5</td>
<td>7</td>
<td>96</td>
<td>.96</td>
</tr>
<tr>
<td>749.5–799.5</td>
<td>4</td>
<td>100</td>
<td>1.00</td>
</tr>
</tbody>
</table>
To form a cumulative frequency polygon, or ogive as it is also called, the cumulative frequency is plotted over the upper boundary of the corresponding class. Figure 14 is the cumulative frequency polygon for the cumulative frequency table in Table 5. Notice that we can easily see that 78% of the students scored below 649.5, while only 18% scored below 499.5. We also can conclude that the probability of a randomly selected score from the sample of 100 lying below 649.5 is .78 and above 649.5 is 1.00 \( - .78 = .22 \).

![Cumulative frequency polygon (ogive)](image)

**FIGURE 14** Cumulative frequency polygon (ogive)

**Insight** Above each class interval in Figure 14, the cumulative frequency polygon is linear. Such a function is said to be piecewise linear. The slope of each piece of a cumulative frequency polygon is greater than or equal to zero (that is, the graph is never falling). The piece with the greatest slope corresponds to the class interval that has the greatest frequency. In fact, for any class interval, the frequency is equal to the slope of the cumulative frequency polygon multiplied by the width of the class interval.

(A) Construct a histogram for the data set whose cumulative frequency polygon is shown in Figure 15.

![Histogram](image)

**FIGURE 15**

(B) Can the original data set be reconstructed from the cumulative frequency polygon? Explain.
Section 8-1  Graphing Data

Answers to Matched Problems

1. (A) .11  (B) .18

2. (A)

(B) .45; .55

Exercise 8-1

1. (A) Construct a frequency table and a histogram for the following data set using a class interval width of 2 starting at 0.5.

6  7  2  7  9
6  4  7  6  6

(B) Construct a frequency table and a histogram for the following data set using a class interval width of 2 starting at 0.5.

5  6  8  1  3
5  10  7  6  8

(C) How are the two histograms of parts (A) and (B) similar? How are the two data sets different?

2. (A) Construct a frequency table and a histogram for the data set of part (A) of Problem 1 using a class interval width of 1 starting at 0.5.

(B) Construct a frequency table and a histogram for the data set of part (B) of Problem 1 using a class interval width of 1 starting at 0.5.

(C) How are the histograms of parts (A) and (B) different?

3. The graphing utility command shown in Figure A generated a set of 400 random integers from 2 to 24, stored as list L1. The statistical plot in Figure B is a histogram of L1, using a class interval width of 1 starting at 1.5.

(A) Explain how the window variables can be changed to display a histogram of the same data set using a class interval width of 2 starting at 1.5. A width of 3 starting at −0.5.

(B) Describe the effect of increasing the class interval width on the shape of the histogram.

4. An experiment consists of rolling a pair of dodecahedral (twelve-sided) dice and recording their sum (the sides of each die are numbered from 1 to 12). The command shown in Figure A simulated 500 rolls of the dodecahedral dice. The statistical plot in Figure B is a histogram of the 500 sums using a class interval width of 1 starting at 1.5.

(A) Explain how the window variables can be changed to display a histogram of the same data set using a class interval width of 2 starting at 1.5. A width of 3 starting at −0.5.

(B) Describe the effect of increasing the class interval width on the shape of the histogram.
Applications

Business & Economics

5. Gross national product. Graph the data in the following table using a bar graph.

<table>
<thead>
<tr>
<th>Year</th>
<th>GNP (billion $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>515.3</td>
</tr>
<tr>
<td>1970</td>
<td>1015.5</td>
</tr>
<tr>
<td>1980</td>
<td>2732.0</td>
</tr>
<tr>
<td>1990</td>
<td>5567.8</td>
</tr>
<tr>
<td>2000</td>
<td>9848.0</td>
</tr>
</tbody>
</table>

6. Corporation revenues. Graph the data in the following table using a bar graph.

<table>
<thead>
<tr>
<th>Corporation</th>
<th>Revenue (million $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wal-Mart</td>
<td>219,812</td>
</tr>
<tr>
<td>Exxon Mobil</td>
<td>191,581</td>
</tr>
<tr>
<td>General Motors</td>
<td>177,260</td>
</tr>
<tr>
<td>Ford Motor</td>
<td>162,412</td>
</tr>
<tr>
<td>Enron</td>
<td>138,718</td>
</tr>
<tr>
<td>General Electric</td>
<td>125,913</td>
</tr>
</tbody>
</table>

7. Gold production. Use the double bar graph on world gold production to determine the country that showed the greatest increase in gold production from 1990 to 2000. Which country showed the greatest percentage increase? Was more gold produced in North America or in South Africa in 1990? In 2000?

8. Gasoline prices. Graph the data in the following table using a divided bar graph.

<table>
<thead>
<tr>
<th>Country</th>
<th>Price Before Tax ($ per gallon)</th>
<th>Tax</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>1.20</td>
<td>.38</td>
</tr>
<tr>
<td>Canada</td>
<td>1.09</td>
<td>.78</td>
</tr>
<tr>
<td>Japan</td>
<td>1.69</td>
<td>2.09</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1.33</td>
<td>3.41</td>
</tr>
<tr>
<td>Germany</td>
<td>1.18</td>
<td>2.53</td>
</tr>
</tbody>
</table>

9. Railroad freight. Graph the data in the following table using a broken-line graph.

<table>
<thead>
<tr>
<th>Year</th>
<th>Carloadings (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>36.4</td>
</tr>
<tr>
<td>1950</td>
<td>38.9</td>
</tr>
<tr>
<td>1960</td>
<td>36.4</td>
</tr>
<tr>
<td>1970</td>
<td>37.1</td>
</tr>
<tr>
<td>1980</td>
<td>27.1</td>
</tr>
<tr>
<td>1990</td>
<td>22.6</td>
</tr>
<tr>
<td>2000</td>
<td>16.4</td>
</tr>
</tbody>
</table>

10. Railroad freight. Refer to Problem 9. If the data were presented in a bar graph, would horizontal bars or vertical bars be used? Could the data be presented in a pie graph? Explain.

11. Federal income. Graph the data in the following table using a pie graph:

<table>
<thead>
<tr>
<th>Source</th>
<th>Income (billion $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal income tax</td>
<td>1,137</td>
</tr>
<tr>
<td>Social insurance taxes</td>
<td>640</td>
</tr>
<tr>
<td>Corporate income tax</td>
<td>236</td>
</tr>
<tr>
<td>Excise tax</td>
<td>55</td>
</tr>
<tr>
<td>Other</td>
<td>30</td>
</tr>
</tbody>
</table>

12. Gasoline prices. In October 2000, the average price of a gallon of gasoline in the United States was $1.532. Of this amount, 75.1 cents was the cost of crude oil, 15.3 cents the cost of refining, 21.4 cents the cost of distribution and marketing, and 41.4 cents the amount of tax. Use a pie graph to present this data.

13. Starting salaries. The starting salaries (in thousands of dollars) of 20 graduates, chosen at random from...
the graduating class of an urban university, were determined and recorded in the table:

<table>
<thead>
<tr>
<th>Starting Salaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>34 29 27 39 41</td>
</tr>
<tr>
<td>28 32 37 35 36</td>
</tr>
<tr>
<td>23 31 33 34 29</td>
</tr>
<tr>
<td>27 35 29 30 32</td>
</tr>
</tbody>
</table>

(A) Construct a frequency and relative frequency table using a class interval of 5 starting at 20.5.
(B) Construct a histogram.
(C) What is the probability that a graduate chosen from the sample will have a starting salary above $32,500? Below $28,500?
(D) Construct a histogram using a graphing utility.

14. Commute times. Thirty-two persons were chosen at random from among the employees of a large corporation and their commute times (in hours) from home to work were determined and recorded in the table:

<table>
<thead>
<tr>
<th>Commute Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 0.9 0.2 0.4 0.7 1.2 1.1 0.7</td>
</tr>
<tr>
<td>0.6 0.4 0.8 1.1 0.9 0.3 0.4 1.0</td>
</tr>
<tr>
<td>0.9 1.0 0.7 0.3 0.6 1.1 0.7 1.1</td>
</tr>
<tr>
<td>0.4 1.3 0.7 0.6 1.0 0.8 0.4 0.9</td>
</tr>
</tbody>
</table>

(A) Construct a frequency and relative frequency table using a class interval width of 0.2 starting at 0.15.
(B) Construct a histogram.
(C) What is the probability that a person chosen at random from the sample will have a commuting time of at least an hour? Of at most half an hour?
(D) Construct a histogram using a graphing utility.

15. Common stocks. The table shows price–earnings ratios of 100 common stocks chosen at random from the New York Stock Exchange.

<table>
<thead>
<tr>
<th>Price–Earnings (PE) Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
<tr>
<td>10 10 10 10 10 10 10 10 10</td>
</tr>
</tbody>
</table>

(A) Construct a frequency and relative frequency table using a class interval of 2 starting at 41.5.
(B) Construct a histogram.
(C) Construct a frequency polygon.
(D) Construct a cumulative frequency and relative cumulative frequency table. What is the probability of a mouse weight drawn at random from the sample lying between 45.5 and 53.5?
(E) Construct a cumulative frequency polygon.

16. Mouse weights. One hundred healthy mice were weighed at the beginning of an experiment with the following results:

<table>
<thead>
<tr>
<th>Mouse Weights (Grams)</th>
</tr>
</thead>
<tbody>
<tr>
<td>51 54 47 53 59 46 50 50 56 46</td>
</tr>
<tr>
<td>48 50 45 49 52 55 42 47 51 45</td>
</tr>
<tr>
<td>53 55 51 47 53 53 49 51 43 48</td>
</tr>
<tr>
<td>44 48 54 46 49 51 52 50 55 51</td>
</tr>
<tr>
<td>50 53 45 49 57 54 53 49 46 48</td>
</tr>
<tr>
<td>52 48 50 52 47 50 44 46 47 49</td>
</tr>
<tr>
<td>49 51 57 49 51 42 49 53 44 52</td>
</tr>
<tr>
<td>53 55 48 52 44 46 54 54 57 55</td>
</tr>
<tr>
<td>48 50 50 55 52 48 47 52 55 50</td>
</tr>
<tr>
<td>59 52 47 46 56 54 51 56 54 55</td>
</tr>
</tbody>
</table>

(A) Construct a frequency and relative frequency table using a class interval of 2 starting at 20.5.
(B) Construct a histogram.
(C) Construct a frequency polygon.
(D) Construct a cumulative frequency and relative cumulative frequency table. What is the probability of a price–earnings ratio drawn at random from the sample lying between 4.5 and 14.5?
(E) Construct a cumulative frequency polygon.

17. Population growth. Graph the data in the following table using a broken-line graph.

<table>
<thead>
<tr>
<th>Annual World Population Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>1900</td>
</tr>
<tr>
<td>1925</td>
</tr>
<tr>
<td>1950</td>
</tr>
<tr>
<td>1975</td>
</tr>
<tr>
<td>2000</td>
</tr>
</tbody>
</table>

18. AIDS epidemic. One way to gauge the toll of the AIDS epidemic in Sub-Saharan Africa is to compare life expectancies with the figures that would have been projected in the absence of AIDS. Use the broken-line graphs shown to estimate the life expectancy of a child born in the year 2002. What
would the life expectancy of the same child be in the absence of AIDS? For which years of birth would the life expectancy be less than 50 years? If there were no AIDS epidemic, for which years of birth would the life expectancy be less than 50 years?

**Life Expectancy in Sub-Saharan Africa**

![Life Expectancy Graph](image)

For which years of birth is the life expectancy less than 50 years?

19. **Nutrition.** Graph the data in the following table using a double bar graph.

<table>
<thead>
<tr>
<th>Grams of:</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 15–18</td>
<td>Age 15–18</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>375</td>
<td>275</td>
</tr>
<tr>
<td>Protein</td>
<td>60</td>
<td>44</td>
</tr>
<tr>
<td>Fat</td>
<td>100</td>
<td>73</td>
</tr>
</tbody>
</table>

20. **Greenhouse gases.** The U.S. Department of Energy estimates that halocarbons account for 20%, methane for 15%, nitrous oxide for 5%, and carbon dioxide for 60% of the enhanced heat-trapping effects of greenhouse gases. Use a pie graph to present this data. Find the central angles of the graph.

21. **Nutrition.** Graph the nutritional information in the following table using a double bar graph.

<table>
<thead>
<tr>
<th>Fast-Food Burgers: Nutritional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calories</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>2-oz burger, plain</td>
</tr>
<tr>
<td>2 addl. oz of beef</td>
</tr>
<tr>
<td>1 slice cheese</td>
</tr>
<tr>
<td>3 slices bacon</td>
</tr>
<tr>
<td>1 tbsp. mayonnaise</td>
</tr>
</tbody>
</table>

22. **Nutrition.** Refer to Problem 21. Suppose that you are trying to limit the fat in your diet to at most 30% of your calories, and your calories to 2000 per day. Should you order the quarter-pound bacon cheeseburger with mayo for lunch? How would such a lunch affect your choice of breakfast and dinner? Discuss.

**Social Sciences**

23. **Education.** In the United States in 1960, 86.4% of school-age children were enrolled in public schools, 12.6% in Catholic schools, and 1.0% in other private schools. In 1998, 86.8% were enrolled in public schools, 4.7% in Catholic schools, 6.5% in other private schools, and 2.0% were home-schooled. Use two pie graphs to present this data.

24. **Study abroad.** Would a pie graph be more effective or less effective than the bar graph shown in presenting information on the most popular destinations of U.S. college students who study abroad? Justify your answer.

**Destinations of U.S. Students Studying Abroad, 1997–1998**

![Pie Chart for Destinations](image)

25. **Median age.** Use the broken-line graph shown to estimate the median age in 1900 and 1990. In which decades did the median age increase? In which did it decrease? Discuss the factors that may have contributed to the increases and decreases.

**Median Age in the United States, 1900–1990**

![Median Age Graph](image)
26. **State prisoners.** In 1980 in the United States, 6% of the inmates of state prisons were incarcerated for drug offenses, 30% for property crimes, 4% for public order offenses, and 59% for violent crimes; in 1998 the percentages were 21%, 21%, 10%, and 48%, respectively. Present the data using two pie graphs. Discuss factors that may account for the shift in percentages between 1980 and 1998.

27. **Grade-point averages.** One hundred seniors were chosen at random from a graduating class at a university and their grade-point averages recorded:

(A) Construct a frequency and relative frequency table using a class interval of 0.2 starting at 1.95.

(B) Construct a histogram.

(C) Construct a frequency polygon.

(D) Construct a cumulative frequency and relative cumulative frequency table. What is the probability of a GPA drawn at random from the sample being over 2.95?

(E) Construct a cumulative frequency polygon.

<table>
<thead>
<tr>
<th>Grade-Point Averages (GPA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 2.0 2.7 2.6 2.1 3.5 3.1 2.1 2.2 2.9</td>
</tr>
<tr>
<td>2.3 2.5 3.1 2.2 2.2 2.0 2.3 2.5 2.1 2.4</td>
</tr>
<tr>
<td>2.7 2.9 2.1 2.2 2.5 2.3 2.1 2.1 3.3 2.1</td>
</tr>
<tr>
<td>2.2 2.2 2.5 2.3 2.7 2.4 2.8 3.1 2.0 2.3</td>
</tr>
<tr>
<td>2.6 3.2 2.2 2.5 3.6 2.3 2.4 3.7 2.5 2.4</td>
</tr>
<tr>
<td>3.5 2.4 2.3 3.9 2.9 2.7 2.6 2.1 2.4 2.0</td>
</tr>
<tr>
<td>2.4 3.3 3.1 2.8 2.3 2.5 2.1 3.0 2.6 2.3</td>
</tr>
<tr>
<td>2.1 2.6 2.2 3.2 2.7 2.8 3.4 2.7 3.6 2.1</td>
</tr>
<tr>
<td>2.7 2.8 3.5 2.4 2.3 2.0 2.1 3.1 2.8 2.1</td>
</tr>
<tr>
<td>3.8 2.5 2.7 2.1 2.2 2.4 2.9 3.3 2.0 2.6</td>
</tr>
</tbody>
</table>

---

**Section 8-2 Measures of Central Tendency**

- **Mean**
- **Median**
- **Mode**

In the preceding section, we found that graphic techniques contributed substantially to our comprehension of large masses of raw data. In this and the next section, we discuss several important numerical measures that are used to describe sets of data. These numerical descriptions are generally of two types:

1. Measures that indicate the approximate center of a distribution, called **measures of central tendency**
2. Measures that indicate the amount of scatter about a central point, called **measures of dispersion**

In this section we look at three widely used measures of central tendency, and in the next section we consider measures of dispersion.

**Mean**

When we speak of the average yield of 10-year municipal bonds, the average number of smog-free days per year in a certain city, or the average SAT score for students at a university, we usually interpret these averages to be **arithmetic averages**, or **means**. In general, we define the mean of a set of quantitative data as follows:

**DEFINITION**

**Mean: Ungrouped Data**

The **mean** of a set of quantitative data is equal to the sum of all the measurements in the data set divided by the total number of measurements in the set.
The mean is a single number that, in a sense, represents the entire data set. It involves all the measurements in the set, it is easily computed, and it enters readily into other useful formulas. Because of these and other desirable properties, the mean is the most widely used measure of central tendency.

In statistics, we are concerned with both a sample mean and the mean of the corresponding population (the sample mean is often used as an estimator for the population mean), so it is important to use different symbols to represent these two means. It is customary to use a letter with an overbar, such as \( \bar{x} \), to represent a sample mean and the Greek letter \( \mu \) (“mu”) to represent a population mean.

**NOTATION**

Mean

\[ \bar{x} = \text{sample mean} \quad \mu = \text{population mean} \]

Before considering examples, let us formulate the concept symbolically using the **summation symbol** \( \Sigma \) (see Appendix B-1). If

\[ x_1, x_2, \ldots, x_n \]

represents a set of \( n \) measurements, then the sum

\[ x_1 + x_2 + \cdots + x_n \]

is compactly and conveniently represented by

\[ \sum_{i=1}^{n} x_i \quad \text{or} \quad \Sigma_{i=1}^{n} x_i \]

We now can express the mean in symbolic form.

**DEFINITION**

Mean: Ungrouped Data

If \( x_1, x_2, \ldots, x_n \) is a set of \( n \) measurements, then the **mean** of the set of measurements is given by

\[ \text{mean} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (1) \]

where

\[ \bar{x} = \text{mean} \text{ if data set is a sample} \]

\[ \mu = \text{mean} \text{ if data set is the population} \]

**Example 1**

**Finding the Mean**

Find the mean for the sample measurements 3, 5, 1, 8, 6, 5, 4, and 6.

**Solution**

Solve using formula (1):

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3 + 5 + 1 + 8 + 6 + 5 + 4 + 6}{8} = \frac{38}{8} = 4.75 \]

**Matched Problem 1**

Find the mean for the sample measurements 3.2, 4.5, 2.8, 5.0, and 3.6.

If data has been grouped in a frequency table, such as Table 4 (Section 8-1), an alternative formula for the mean is generally used:
DEFINITION

Mean: Grouped Data

A data set of \( n \) measurements is grouped into \( k \) classes in a frequency table. If \( x_i \) is the midpoint of the \( i \)th class interval and \( f_i \) is the \( i \)th class frequency, the mean for the grouped data is given by

\[
[\text{mean}] = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{n}
\]

where

\[
n = \sum_{i=1}^{k} f_i = \text{total number of measurements}
\]

\[
\bar{x} = [\text{mean}] \text{ if data set is a sample}
\]

\[
\mu = [\text{mean}] \text{ if data set is the population}
\]

Caution

It is important to note that \( n \) is the total number of measurements in the entire data set—not the number of classes!

The mean computed by formula (2) is a weighted average of the midpoints of the class intervals. In general, this will be close to, but not exactly the same as, the mean computed by formula (1) for ungrouped data.

EXAMPLE 2

Finding the Mean for Grouped Data

Find the mean for the sample data summarized in Table 4, Section 8-1.

Solution

We repeat part of Table 4 here, adding columns for the class midpoints \( x_i \) and the products \( x_i f_i \) (Table 1).

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Midpoint ( x_i )</th>
<th>Frequency ( f_i )</th>
<th>Product ( x_i f_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>299.5–349.5</td>
<td>324.5</td>
<td>1</td>
<td>324.5</td>
</tr>
<tr>
<td>349.5–399.5</td>
<td>374.5</td>
<td>2</td>
<td>749.0</td>
</tr>
<tr>
<td>399.5–449.5</td>
<td>424.5</td>
<td>5</td>
<td>2,122.5</td>
</tr>
<tr>
<td>449.5–499.5</td>
<td>474.5</td>
<td>10</td>
<td>4,745.0</td>
</tr>
<tr>
<td>499.5–549.5</td>
<td>524.5</td>
<td>21</td>
<td>11,014.5</td>
</tr>
<tr>
<td>549.5–599.5</td>
<td>574.5</td>
<td>20</td>
<td>11,490.0</td>
</tr>
<tr>
<td>599.5–649.5</td>
<td>624.5</td>
<td>19</td>
<td>11,865.5</td>
</tr>
<tr>
<td>649.5–699.5</td>
<td>674.5</td>
<td>11</td>
<td>7,419.5</td>
</tr>
<tr>
<td>699.5–749.5</td>
<td>724.5</td>
<td>7</td>
<td>5,071.5</td>
</tr>
<tr>
<td>749.5–799.5</td>
<td>774.5</td>
<td>4</td>
<td>3,098.0</td>
</tr>
</tbody>
</table>

\[
n = \sum_{i=1}^{10} f_i = 100
\]

\[
\sum_{i=1}^{10} x_i f_i = 57,900.0
\]

Thus, the average entrance examination score for the sample of 100 entering freshmen is

\[
\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{57,900}{100} = 579
\]

If the histogram for the data in Table 1 (Fig. 11, Section 8-1) was drawn on a piece of wood of uniform thickness and the wood cut around the outside of the figure, the resulting object would balance exactly at the mean \( \bar{x} = 579 \), as shown in Figure 1.
Compute the mean for the grouped sample data listed in Table 2.

**TABLE 2**

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–5.5</td>
<td>6</td>
</tr>
<tr>
<td>5.5–10.5</td>
<td>20</td>
</tr>
<tr>
<td>10.5–15.5</td>
<td>18</td>
</tr>
<tr>
<td>15.5–20.5</td>
<td>4</td>
</tr>
</tbody>
</table>

**Insight**

The mean for ungrouped data and the mean for grouped data can be interpreted as the expected values of appropriately chosen random variables (see Section 7-5).

Consider a set of \( n \) measurements \( x_1, x_2, \ldots, x_n \) (ungrouped data). Let \( S \) be the sample space consisting of \( n \) simple events (the \( n \) measurements), each equally likely. Let \( X \) be the random variable that assigns the numerical value \( x_i \) to each simple event in \( S \). Then each measurement \( x_i \) has probability \( p_i = \frac{1}{n} \). The expected value of \( X \) is given by

\[
E(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n
\]

\[
= x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + \cdots + x_n \cdot \frac{1}{n}
\]

\[
= x_1 + x_2 + \cdots + x_n
\]

\[
= \text{[mean]}
\]

Similarly, consider a set of \( n \) measurements grouped into \( k \) classes in a frequency table (grouped data). Let \( S' \) be the sample space consisting of \( n \) simple events (the \( n \) measurements), each equally likely. Let \( X' \) be the random variable that assigns the midpoint \( x_i \) of the \( i \)th class interval to the measurements that belong to that class interval. Then each midpoint \( x_i \) has probability \( p_i = \frac{f_i}{n} \), where \( f_i \) denotes the frequency of the \( i \)th class interval. The expected value of \( X' \) is given by

\[
E(X') = x_1p_1 + x_2p_2 + \cdots + x_kp_k
\]

\[
= x_1\left(\frac{f_1}{n}\right) + x_2\left(\frac{f_2}{n}\right) + \cdots + x_k\left(\frac{f_k}{n}\right)
\]

\[
= \frac{x_1f_1 + x_2f_2 + \cdots + x_kf_k}{n}
\]

\[
= \text{[mean]}
\]

**Median**

Occasionally, the mean can be misleading as a measure of central tendency. Suppose the annual salaries of seven people in a small company are $17,000, $20,000, $28,000, $18,000, $18,000, $120,000, and $24,000. The mean salary is

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{$245,000}{7} = $35,000
\]
Six of the seven salaries are below the average! The one large salary distorts the results.

A measure of central tendency that is not influenced by extreme values is the **median**. The following definition of median makes precise our intuitive notion of the “middle element” when a set of measurements is arranged in ascending or descending order. Some sets of measurements, for example, 5, 7, 8, 13, 21, have a middle element. Other sets, for example, 9, 10, 15, 20, 23, 24, have no middle element, or you might prefer to say they have two middle elements. For any number between 15 and 20, half the measurements fall above the number and half fall below.

**Definition**

**Median**

1. If the number of measurements in a set is odd, the **median** is the middle measurement when the measurements are arranged in ascending or descending order.

2. If the number of measurements in a set is even, the **median** is the mean of the two middle measurements when the measurements are arranged in ascending or descending order.

**Example 3**

**Finding the Median** Find the median salary in the preceding list of seven salaries.

**Solution**

Arrange the salaries in increasing order and choose the middle one:

<table>
<thead>
<tr>
<th>SALARY</th>
<th></th>
<th></th>
<th></th>
<th>Median ($20,000)</th>
<th></th>
<th>Mean ($35,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$17,000</td>
<td>18,000</td>
<td>18,000</td>
<td></td>
<td>20,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20,000</td>
<td>24,000</td>
<td>28,000</td>
<td></td>
<td>120,000</td>
<td></td>
</tr>
</tbody>
</table>

In this case, the median is a better measure of central tendency than the mean.

**Matched Problem 3** Add the salary $100,000 to those in Example 3 and compute the median and mean for these eight salaries.

The median, as we have defined it, is easy to determine and is not influenced by extreme values. Our definition does have some minor handicaps, however. First, if the measurements we are analyzing were carried out in a laboratory and presented to us in a frequency table, we may not have access to the individual measurements. In that case we would not be able to compute the median using the above definition. Second, a set like 4, 4, 6, 7, 7, 7, 9 would have median 7 by our definition, but 7 does not possess the symmetry we expect of a “middle element” since there are three measurements below 7 but only one above.

To overcome these handicaps, we define a second concept, the **median for grouped data**. To guarantee that the median for grouped data exists and is unique, we assume that the frequency table for the grouped data has no classes of frequency 0.
DEFINITION  
Median for Grouped Data
The median for grouped data with no classes of frequency 0 is the number such that the histogram has the same area to the left of the median as to the right of the median (see Fig. 2).

FIGURE 2  The area to the left of the median equals the area to the right.

EXAMPLE 4  Finding the Median for Grouped Data  Compute the median for the grouped data of Table 3.

Solution
We first draw the histogram of the data (Fig. 3). The total area of the histogram is 15, which is just the sum of the frequencies, since all rectangles have a base of length 1. The area to the left of the median must be half the total area—that is, \( \frac{15}{2} = 7.5 \). Looking at Figure 3 we see that the median \( M \) lies between 6.5 and 7.5. Thus, the area to the left of \( M \), which is the sum of the blue shaded areas in Figure 3, must be 7.5:

\[
(1)(3) + (1)(1) + (1)(2) + (M - 6.5)(4) = 7.5
\]

Solving for \( M \) gives \( M = 6.875 \). That is, the median for the grouped data in Table 3 is 6.875.

Matched Problem 4  Find the median for the grouped data in the following table:

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5–4.5</td>
<td>4</td>
</tr>
<tr>
<td>4.5–5.5</td>
<td>2</td>
</tr>
<tr>
<td>5.5–6.5</td>
<td>3</td>
</tr>
<tr>
<td>6.5–7.5</td>
<td>5</td>
</tr>
<tr>
<td>7.5–8.5</td>
<td>4</td>
</tr>
<tr>
<td>8.5–9.5</td>
<td>3</td>
</tr>
</tbody>
</table>
We have given geometric interpretations of the mean for grouped data as the balance point of a histogram (Fig. 1), and of the median for grouped data as the point that divides a histogram into equal areas (Fig. 2).

(A) Give an example of a simple histogram in which the balance point (the mean) and the point marking equal areas (the median) are two different points. Explain how other examples in which the mean does not equal the median could be constructed.

(B) Give an example of a simple histogram in which the balance point (the mean) and the point marking equal areas (the median) are the same point. Discuss properties of a histogram that would guarantee that the mean and the median are equal.

(C) Explain why the median for grouped data is not influenced by extreme values.

**Mode**

A third measure of central tendency is the *mode*.

**DEFINITION**

*Mode*

The mode is the most frequently occurring measurement in a data set. There may be a unique mode, several modes, or, if no measurement occurs more than once, essentially no mode.

**Example 5**

Finding Mode, Median, and Mean

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Mode</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) 4, 5, 5, 5, 6, 6, 7, 8, 12</td>
<td>5</td>
<td>6</td>
<td>6.44</td>
</tr>
<tr>
<td>(B) 1, 2, 3, 3, 3, 5, 6, 7, 7, 23</td>
<td>3, 7</td>
<td>5</td>
<td>6.09</td>
</tr>
<tr>
<td>(C) 1, 3, 5, 6, 7, 9, 11, 15, 16</td>
<td>None</td>
<td>7</td>
<td>8.11</td>
</tr>
</tbody>
</table>

Data set (B) in Example 5 is referred to as *bimodal*, since there are two modes. Since no measurement in data set (C) occurs more than once, we say that it has no mode.

**Matched Problem 5**

Compute the mode(s), median, and mean for each data set:

(A) 2, 1, 2, 1, 1, 5, 1, 9, 4   (B) 2, 5, 1, 4, 9, 8, 7   (C) 8, 2, 6, 8, 3, 3, 1, 5, 1, 8, 3

The mode, median, and mean can be computed in various ways with the aid of a graphing utility. In Figure 4A the data set of Example 5B is entered as a
list, and its median and mean are computed. The histogram in Figure 4B shows the two modes of the same data set.

As with the median, the mode is not influenced by extreme values. Suppose, in the data set of Example 5B, we replace 23 by 8. The modes remain 3 and 7 and the median is still 5, but the mean changes to 4.73. The mode is most useful for large data sets because it emphasizes data concentration. For example, a clothing retailer would be interested in the mode of sizes due to customer demand of the various items stocked in a store.

The mode also can be used for qualitative attributes—that is, attributes that are not numerical. The mean and median are not suitable in these cases. For example, the mode can be used to give an indication of a favorite brand of ice cream or the worst movie of the year. Figure 5 shows the results of a random survey of 1,000 people on entree preferences when eating dinner out. According to this survey, we would say that the modal preference is beef. Note that the mode is the only measure of central tendency (location) that can be used for this type of data; the mean and median make no sense.

In actual practice, the mean is used the most, the median next, and the mode a distant third.

For many sets of measurements the median lies between the mode and the mean. But this is not always so.

(A) In a class of seven students the scores on an exam were 52, 89, 89, 92, 93, 96, 99. Show that the mean is less than the mode, and that the mode is less than the median.

(B) Construct hypothetical sets of exam scores to show that all possible orders among the mean, median, and mode can occur.

**Answers to Matched Problems**

1. $\bar{x} \approx 3.8$
2. $\bar{x} \approx 10.1$
3. Median = $22,000$; mean = $43,125$
4. Median for grouped data = 6.8
5. First, arrange each set of data in ascending order:

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Mode</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) 1, 1, 1, 2, 2, 4, 5, 9</td>
<td>1</td>
<td>2</td>
<td>2.89</td>
</tr>
<tr>
<td>(B) 1, 2, 4, 5, 7, 8, 9</td>
<td>None</td>
<td>5</td>
<td>5.14</td>
</tr>
<tr>
<td>(C) 1, 1, 2, 3, 3, 5, 6, 8, 8, 8</td>
<td>3, 8</td>
<td>3</td>
<td>4.36</td>
</tr>
</tbody>
</table>

Exercise 8-2

A Find the mean, median, and mode for the sets of ungrouped data given in Problems 1 and 2.

1. 1, 2, 2, 3, 3, 3, 4, 4, 5

2. 1, 1, 1, 2, 3, 4, 5, 5, 5

Find the mean, median, and/or mode, whichever are applicable, in Problems 3 and 4.

3. Flavor | Number Preferring
--- | ---
Vanilla | 139
Chocolate | 376
Strawberry | 89
Pistachio | 105
Cherry | 63
Almond mocha | 228

4. Car Color | Number Preferring
--- | ---
Red | 1,324
White | 3,084
Black | 1,617
Blue | 2,303
Brown | 2,718
Gold | 1,992

Find the mean for the sets of grouped data in Problems 5 and 6.

5. Interval | Frequency
--- | ---
0.5–2.5 | 2
2.5–4.5 | 5
4.5–6.5 | 7
6.5–8.5 | 1

6. Interval | Frequency
--- | ---
0.5–2.5 | 5
2.5–4.5 | 1
4.5–6.5 | 2
6.5–8.5 | 7

B 7. Which single measure of central tendency—the mean, median, or mode—would you say best describes the following set of measurements? Discuss the factors that justify your preference.

| 8.01 | 7.91 | 8.13 | 6.24 | 7.95 |
| 8.04 | 7.99 | 8.09 | 6.24 | 81.2 |

8. Which single measure of central tendency—the mean, median, or mode—would you say best describes the following set of measurements? Discuss the factors that justify your preference.

| 47 | 51 | 80 | 91 | 85 |
| 69 | 91 | 95 | 81 | 60 |

9. A data set is formed by recording the results of 100 rolls of a fair die.

(A) What would you expect the mean of the data set to be? The median?

(B) Form such a data set by using a graphing utility to simulate 100 rolls of a fair die, and find its mean and median.

10. A data set is formed by recording the sums on 200 rolls of a pair of fair dice.

(A) What would you expect the mean of the data set to be? The median?

(B) Form such a data set by using a graphing utility to simulate 200 rolls of a pair of fair dice, and find the mean and median of the set.

C 11. (A) Construct a set of four numbers that has mean 300, median 250, and mode 175.

(B) Let $m_1 > m_2 > m_3$. Devise and discuss a procedure for constructing a set of four numbers that has mean $m_1$, median $m_2$, and mode $m_3$.

12. (A) Construct a set of five numbers that has mean 200, median 150, and mode 50.

(B) Let $m_1 > m_2 > m_3$. Devise and discuss a procedure for constructing a set of five numbers that has mean $m_1$, median $m_2$, and mode $m_3$. 


### Applications

#### Business & Economics

13. **Price–earnings ratios.** Find the mean, median, and mode for the data in the following table.

<table>
<thead>
<tr>
<th>Price–Earnings Ratios for Eight Stocks in a Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3</td>
</tr>
<tr>
<td>12.9</td>
</tr>
</tbody>
</table>

14. **Gasoline tax.** Find the mean, median, and mode for the data in the following table.

<table>
<thead>
<tr>
<th>State Gasoline Tax, 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Wisconsin</td>
</tr>
<tr>
<td>New York</td>
</tr>
<tr>
<td>Connecticut</td>
</tr>
<tr>
<td>Nebraska</td>
</tr>
<tr>
<td>Kansas</td>
</tr>
<tr>
<td>Texas</td>
</tr>
<tr>
<td>California</td>
</tr>
<tr>
<td>Florida</td>
</tr>
</tbody>
</table>

15. **Light bulb lifetime.** Find the mean and median for the data in the following table.

<table>
<thead>
<tr>
<th>Life (Hours) of 50 Randomly Selected Light Bulbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>799.5–899.5</td>
</tr>
<tr>
<td>899.5–999.5</td>
</tr>
<tr>
<td>999.5–1,099.5</td>
</tr>
<tr>
<td>1,099.5–1,199.5</td>
</tr>
<tr>
<td>1,199.5–1,299.5</td>
</tr>
</tbody>
</table>

16. **Price–earnings ratios.** Find the mean and median for the data in the following table.

<table>
<thead>
<tr>
<th>Price–Earnings Ratios of 100 Randomly Chosen Stocks from The New York Stock Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>−0.5–4.5</td>
</tr>
<tr>
<td>4.5–9.5</td>
</tr>
<tr>
<td>9.5–14.5</td>
</tr>
<tr>
<td>14.5–19.5</td>
</tr>
<tr>
<td>19.5–24.5</td>
</tr>
<tr>
<td>24.5–29.5</td>
</tr>
<tr>
<td>29.5–34.5</td>
</tr>
</tbody>
</table>

#### Financial aid

17. **Financial aid.** Find the mean, median, and mode for the data on federal student financial assistance in the following table.

<table>
<thead>
<tr>
<th>Average Federal Work–Study Award</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1995</td>
</tr>
<tr>
<td>1996</td>
</tr>
<tr>
<td>1997</td>
</tr>
<tr>
<td>1998</td>
</tr>
<tr>
<td>1999</td>
</tr>
<tr>
<td>2000</td>
</tr>
<tr>
<td>2001</td>
</tr>
</tbody>
</table>

#### Tourism

18. **Tourism.** Find the mean, median, and mode for the data in the following table.

<table>
<thead>
<tr>
<th>International Tourism Receipts, 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>United States</td>
</tr>
<tr>
<td>Spain</td>
</tr>
<tr>
<td>France</td>
</tr>
<tr>
<td>Italy</td>
</tr>
<tr>
<td>China</td>
</tr>
<tr>
<td>Germany</td>
</tr>
<tr>
<td>United Kingdom</td>
</tr>
<tr>
<td>Austria</td>
</tr>
<tr>
<td>Canada</td>
</tr>
<tr>
<td>Greece</td>
</tr>
</tbody>
</table>

#### Life Sciences

19. **Mouse weights.** Find the mean and median for the data in the following table.

<table>
<thead>
<tr>
<th>Mouse Weights (Grams)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>41.5–43.5</td>
</tr>
<tr>
<td>43.5–45.5</td>
</tr>
<tr>
<td>45.5–47.5</td>
</tr>
<tr>
<td>47.5–49.5</td>
</tr>
<tr>
<td>49.5–51.5</td>
</tr>
<tr>
<td>51.5–53.5</td>
</tr>
<tr>
<td>53.5–55.5</td>
</tr>
<tr>
<td>55.5–57.5</td>
</tr>
<tr>
<td>57.5–59.5</td>
</tr>
</tbody>
</table>
20. **Blood cholesterol levels.** Find the mean and median for the data in the following table:

<table>
<thead>
<tr>
<th>Blood Cholesterol Levels (Milligrams per Deciliter)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>149.5–169.5</td>
<td>4</td>
</tr>
<tr>
<td>169.5–189.5</td>
<td>11</td>
</tr>
<tr>
<td>189.5–209.5</td>
<td>15</td>
</tr>
<tr>
<td>209.5–229.5</td>
<td>25</td>
</tr>
<tr>
<td>229.5–249.5</td>
<td>13</td>
</tr>
<tr>
<td>249.5–269.5</td>
<td>7</td>
</tr>
<tr>
<td>269.5–289.5</td>
<td>3</td>
</tr>
<tr>
<td>289.5–309.5</td>
<td>2</td>
</tr>
</tbody>
</table>

**Social Sciences**

21. **Immigration.** Find the mean, median, and mode for the data in the following table:

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
</tr>
<tr>
<td>Mexico</td>
</tr>
<tr>
<td>Philippines</td>
</tr>
<tr>
<td>China</td>
</tr>
<tr>
<td>India</td>
</tr>
<tr>
<td>Cuba</td>
</tr>
<tr>
<td>Vietnam</td>
</tr>
<tr>
<td>El Salvador</td>
</tr>
<tr>
<td>Korea</td>
</tr>
<tr>
<td>Dominican Republic</td>
</tr>
<tr>
<td>Great Britain</td>
</tr>
</tbody>
</table>

22. **Grade-point averages.** Find the mean and median for the grouped data in the following table:

<table>
<thead>
<tr>
<th>Graduating Class Grade-Point Averages</th>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.95–2.15</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>2.15–2.35</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>2.35–2.55</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2.55–2.75</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.75–2.95</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2.95–3.15</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>3.15–3.35</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3.35–3.55</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3.55–3.75</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3.75–3.95</td>
<td>2</td>
</tr>
</tbody>
</table>

23. **Entrance examination scores.** Compute the median for the grouped data of entrance examination scores given in Table 1.

24. **Presidents.** Find the mean and median for the grouped data in the following table:

<table>
<thead>
<tr>
<th>U.S. Presidents’ Ages at Inauguration</th>
<th>Age</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39.5–44.5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>44.5–49.5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>49.5–54.5</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>54.5–59.5</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>59.5–64.5</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>64.5–69.5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>69.5–74.5</td>
<td>1</td>
</tr>
</tbody>
</table>

---

**Section 8-3 Measures of Dispersion**

- **Range**
- **Standard Deviation: Ungrouped Data**
- **Standard Deviation: Grouped Data**
- **Significance of Standard Deviation**

A measure of central tendency gives us a typical value that can be used to describe a whole set of data, but this measure does not tell us whether the data are tightly clustered or widely dispersed. We now consider two measures of variation—range and standard deviation—that will give some indication of data scatter.

**Range**

A measure of dispersion, or scatter, that is easy to compute and is easily understood is the range. The *range for a set of ungrouped data* is the difference between the largest and the smallest values in the data set. The *range for a frequency distribution* is the difference between the upper boundary of the highest class and the lower boundary of the lowest class.
Consider the histograms in Figure 1. We see that the range adds only a little information about the amount of variation in a data set. The graphs clearly show that even though each data set has the same mean and range, all three sets differ in the amount of scatter, or variation, of the data relative to the mean. The data set in part (A) is tightly clustered about the mean; the data set in part (B) is dispersed away from the mean; and the data set in part (C) is uniformly distributed over its range.

Since the range depends only on the extreme values of the data, it does not give us any information about the dispersion of the data between these extremes. We need a measure of dispersion that will give us some idea of how the data are clustered or scattered relative to the mean. The standard deviation is such a measure.

**Standard Deviation: Ungrouped Data**

We will develop the concepts of variance and standard deviation—both measures of variation—through a simple example. Suppose that a random sample of five stamped parts is selected from a manufacturing process, and these parts are found to have the following lengths (in centimeters):

5.2, 5.3, 5.2, 5.5, 5.3

Computing the sample mean, we obtain

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5.2 + 5.3 + 5.2 + 5.5 + 5.3}{5} = 5.3 \text{ centimeters} \]

How much variation exists between the sample mean and all measurements in the sample? As a first attempt at measuring the variation, let us represent the deviation of a measurement from the mean by \((x_i - \bar{x})\). Table 1 lists all the deviations for this sample.

Using these deviations, what kind of formula can we find that will give us a single measure of variation? It appears that the average of the deviations might be a good measure. But look what happens when we add the second column in Table 1. We get 0! It turns out that this will always happen for any data set. Now what? We could take the average of the absolute values of the deviations; however, this approach leads to problems relative to statistical inference. Instead, to get around the sign problem, we will take the average of the squares of the deviations and call this number the variance of the data set:

\[
[\text{variance}] = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \tag{1}
\]

Calculating the variance using the entries in Table 1, we have

\[ [\text{variance}] = \frac{\sum_{i=1}^{5} (x_i - 5.3)^2}{5} = 0.012 \text{ square centimeter} \]

We still have a problem, because the units in the variance are square centimeters instead of centimeters (the units of the original data set). To obtain the units of the original data set, we take the positive square root of the variance.
and call the result the **standard deviation** of the data set:

$$[\text{standard deviation}] = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{5} (x_i - 5.3)^2}{5}}$$  (2)

$$= 0.11 \text{ centimeter}$$

The sample variance is usually denoted by $s^2$ and the population variance by $\sigma^2$ ($\sigma$ is the Greek lowercase letter “sigma”). The sample standard deviation is usually denoted by $s$ and the population standard deviation by $\sigma$.

In inferential statistics the sample variance $s^2$ is often used as an estimator for the population variance $\sigma^2$ and the sample standard deviation $s$ for the population standard deviation $\sigma$. It can be shown that one can obtain better estimates of the population parameters in terms of the sample parameters (particularly when using small samples) if the divisor $n$ is replaced by $n - 1$ when computing sample variances or sample standard deviations. With these modifications, we have the following:

---

**DEFINITION** Variance: Ungrouped Data*

The **sample variance** $s^2$ of a set of $n$ sample measurements $x_1, x_2, \ldots, x_n$ with mean $\bar{x}$ is given by

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$  (3)

If $x_1, x_2, \ldots, x_n$ is the whole population with mean $\mu$, then the **population variance** $\sigma^2$ is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

* In this section we restrict our interest to the sample variance.

---

The standard deviation is just the positive square root of the variance. Therefore, we have the following formulas:

**DEFINITION** Standard Deviation: Ungrouped Data*

The **sample standard deviation** $s$ of a set of $n$ sample measurements $x_1, x_2, \ldots, x_n$ with mean $\bar{x}$ is given by

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$  (4)

If $x_1, x_2, \ldots, x_n$ is the whole population with mean $\mu$, then the **population standard deviation** $\sigma$ is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

* In this section we restrict our interest to the sample standard deviation.
Computing the standard deviation for the original sample measurements (Table 1), we now obtain

\[
s = \sqrt{\sum_{i=1}^{5} (x_i - 5.3)^2} = 0.12 \text{ centimeter}
\]

**Example 1**

**Finding the Standard Deviation** Find the standard deviation for the sample measurements 1, 3, 5, 4, 3.

**Solution**

To find the standard deviation for the data set, we can utilize a table or use a calculator. Most will prefer the latter. Here is what we compute:

\[
\bar{x} = \frac{1 + 3 + 5 + 4 + 3}{5} = 3.2
\]

\[
s = \sqrt{\frac{(1 - 3.2)^2 + (3 - 3.2)^2 + (5 - 3.2)^2 + (4 - 3.2)^2 + (3 - 3.2)^2}{5 - 1}}
\]

\[
\approx 1.48
\]

**Matched Problem 1**

Find the standard deviation for the sample measurements 1.2, 1.4, 1.7, 1.3, 1.5.

**Remark**

Many calculators and graphing utilities can compute \(\bar{x}\) and \(s\) directly after the sample measurements are entered—a helpful feature, especially when the sample is fairly large. This shortcut is illustrated in Figure 2 for a particular graphing calculator, where the data from Example 1 are entered as a list and several different one-variable statistics are immediately calculated. Included among these statistics are the mean \(\bar{x}\), the sample standard deviation \(s\) (denoted by Sx in Fig. 2B), the population standard deviation \(\sigma\) (denoted by \(\sigma x\)), the number \(n\) of measurements, the smallest element of the data set (denoted by minX), the largest element of the data set (denoted by maxX), the median (denoted by Med), and several statistics we have not discussed.

![Figure 2](image)

**Insight**

If the sample measurements in Example 1 are considered to constitute the whole population, then the population standard deviation \(\sigma x\) is approximately equal to 1.33 [see Fig. 2(B)]. The computation of \(\sigma x\) is the same as that of Example 1, except that the denominator \(n - 1 (= 5 - 1)\) under the radical sign is replaced by \(n (= 5)\). Consequently, Sx \(\approx 1.48\) is greater than \(\sigma x \approx 1.33\). Formulas (4) and (2) produce nearly the same results when the sample size \(n\) is large. The law of large numbers states that we can make a
sample standard deviation \( s \) as close to the population standard deviation \( \sigma \) as we like by making the sample sufficiently large.

**Explore & Discuss 1**

(A) When is the sample standard deviation of a set of measurements equal to 0?
(B) Can the sample standard deviation of a set of measurements ever be greater than the range? Explain why or why not.

---

**Standard Deviation: Grouped Data**

Formula (4) for sample standard deviation is extended to grouped sample data as described in the following box:

**Definition**

Suppose a data set of \( n \) sample measurements is grouped into \( k \) classes in a frequency table, where \( x_i \) is the midpoint and \( f_i \) is the frequency of the \( i \)th class interval. Then the **sample standard deviation** \( s \) for the grouped data is

\[
s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n - 1}}
\]

where \( n = \sum_{i=1}^{k} f_i \) = total number of measurements. If \( x_1, x_2, \ldots, x_n \) is the whole population with mean \( \mu \), then the **population standard deviation** \( \sigma \) is given by

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2 f_i}{n}}
\]

*In this section we restrict our interest to the sample standard deviation.*

**Example 2**

**Finding the Standard Deviation for Grouped Data**

Find the standard deviation for each set of grouped sample data.

(A) \( s = \sqrt{\frac{(8 - 10)^2(1) + (9 - 10)^2(2) + (10 - 10)^2(4) + (11 - 10)^2(2) + (12 - 10)^2(1)}{10 - 1}} \)

\[
= \sqrt{\frac{12}{9}} \approx 1.15
\]

(B) \( s = \sqrt{\frac{(8 - 10)^2(4) + (9 - 10)^2(1) + (10 - 10)^2(0) + (11 - 10)^2(1) + (12 - 10)^2(4)}{10 - 1}} \)

\[
= \sqrt{\frac{34}{9}} \approx 1.94
\]
Comparing the results of parts (A) and (B) in Example 2, we find that the larger standard deviation is associated with the data that deviate furthest from the mean.

Find the standard deviation for the grouped sample data shown below.

<table>
<thead>
<tr>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Mean = 10

Remark

Figure 3 illustrates the shortcut computation of the mean and standard deviation on a particular graphing calculator when the data are grouped. The sample data of Example 2A are entered. List \( L_1 \) contains the midpoints of the class intervals, and list \( L_2 \) contains the corresponding frequencies. The mean, standard deviation, and other one-variable statistics are then calculated immediately.

A smooth curve can be drawn through the midpoints of the tops of the rectangles forming a histogram for a fairly large frequency distribution (see Fig. 4). If the resulting curve is approximately bell-shaped, then it can be shown that approximately 68% of the data will lie in the interval from \( \bar{x} - s \) to \( \bar{x} + s \), about 95% of the data will lie in the interval from \( \bar{x} - 2s \) to \( \bar{x} + 2s \), and almost all the data will lie in the interval from \( \bar{x} - 3s \) to \( \bar{x} + 3s \). We will have much more to say about this in Section 8-5.
(A) Verify that the following sample of 21 measurements has mean \( \bar{x} = 4.95 \) and sample standard deviation \( s = 2.31 \). Therefore, 15 of the 21 measurements, or 71% of the data, lie in the interval 2.64–7.26, that is, within 1 standard deviation of the mean. What proportion of the data lies within 2 standard deviations of the mean? Within 3 standard deviations?

\[
\begin{align*}
4 & 5 & 9 & 3 & 4 & 6 & 1 \\
5 & 0 & 7 & 2 & 5 & 8 & 6 \\
3 & 4 & 7 & 6 & 5 & 6 & 8 \\
\end{align*}
\]

(B) What proportion of the following data set lies within 1 standard deviation of the mean? Within 2 standard deviations? Within 3 standard deviations?

\[
\begin{align*}
2 & 8 & 0 & 9 & 2 & 8 & 1 \\
3 & 5 & 6 & 0 & 1 & 9 & 4 \\
5 & 8 & 2 & 1 & 3 & 7 & 9 \\
\end{align*}
\]

(C) Based on your answers to parts (A) and (B), which of the two data sets would have a histogram that is approximately bell shaped? Confirm your conjecture by constructing a histogram with class interval width 1, starting at –0.5, for each data set.

**Answers to Matched Problems**

1. \( s \approx 0.19 \)
2. \( s \approx 1.49 \) (a value between those found in Example 2, as expected)

**Exercise 8-3**

**A** In Problems 1 and 2, find the standard deviation for each set of ungrouped sample data using formula (4).

1. 1, 2, 2, 3, 3, 3, 3, 4, 4, 5
2. 1, 1, 1, 1, 2, 3, 4, 5, 5, 5

3. (A) What proportion of the following sample of ten measurements lies within 1 standard deviation of the mean? Within 2 standard deviations? Within 3 standard deviations?

\[
\begin{align*}
4 & 2 & 3 & 5 & 3 \\
1 & 6 & 4 & 2 & 3 \\
\end{align*}
\]

(B) Based on your answers to part (A), would you conjecture that the histogram is approximately bell shaped? Explain.

(C) To confirm your conjecture, construct a histogram with class interval width 1, starting at 0.5.

4. (A) What proportion of the following sample of ten measurements lies within 1 standard deviation of the mean? Within 2 standard deviations? Within 3 standard deviations?

\[
\begin{align*}
3 & 5 & 1 & 2 & 1 \\
5 & 4 & 5 & 1 & 3 \\
\end{align*}
\]

(B) Based on your answers to part (A), would you conjecture that the histogram is approximately bell shaped? Explain.

(C) To confirm your conjecture, construct a histogram with class interval width 1, starting at 0.5.

**B** In Problems 5 and 6, find the standard deviation for each set of grouped sample data using formula (5).

5.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–3.5</td>
<td>2</td>
</tr>
<tr>
<td>3.5–6.5</td>
<td>5</td>
</tr>
<tr>
<td>6.5–9.5</td>
<td>7</td>
</tr>
<tr>
<td>9.5–12.5</td>
<td>1</td>
</tr>
</tbody>
</table>

6.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–3.5</td>
<td>5</td>
</tr>
<tr>
<td>3.5–6.5</td>
<td>1</td>
</tr>
<tr>
<td>6.5–9.5</td>
<td>2</td>
</tr>
<tr>
<td>9.5–12.5</td>
<td>7</td>
</tr>
</tbody>
</table>
In Problems 7 and 8, discuss the validity of each statement. If the statement is always true, explain why. If not, give a counterexample.

7. (A) The sample variance of a set of \( n \) sample measurements is always greater than or equal to the sample standard deviation.
(B) The population variance of \( x_1, x_2, \ldots, x_n \) is always greater than or equal to 0.

8. (A) The sample variance of a set of \( n \) sample measurements is always positive.
(B) For a sample \( x_1, x_2 \) of size two, the sample variance is equal to 
\[
\frac{(x_1 - x_2)^2}{2}
\]

9. A data set is formed by recording the sums in 100 rolls of a pair of dice. A second data set is formed by recording the results of 100 draws of a ball from a box containing 11 balls numbered 2 through 12.

(A) Which of the two data sets would you expect to have the smaller standard deviation? Explain.
(B) To obtain evidence for your answer to part (A), use a graphing utility to simulate both experiments, and compute the standard deviations of each of the two data sets.

10. A data set is formed by recording the results of rolling a fair die 200 times. A second data set is formed by rolling a pair of dice 200 times, each time recording the minimum of the two numbers.

(A) Which of the two data sets would you expect to have the smaller standard deviation? Explain.
(B) To obtain evidence for your answer to part (A), use a graphing utility to simulate both experiments, and compute the standard deviations of each of the two data sets.

Applications

Find the mean and standard deviation for each of the sample data sets given in Problems 11–18.

Use the suggestions in the remarks following Examples 1 and 2 to perform some of the computations.

Business & Economics

11. Earnings per share. The earnings per share (in dollars) for 12 companies selected at random from the list of Fortune 500 companies are

\[
\begin{array}{cccc}
2.35 & 3.11 & 5.33 \\
1.42 & 2.56 & 7.74 \\
8.05 & 0.72 & 3.88 \\
6.71 & 4.17 & 6.21
\end{array}
\]

12. Checkout times. The checkout times (in minutes) for 12 randomly selected customers at a large supermarket during the store’s busiest time are

\[
\begin{array}{cccc}
4.6 & 8.5 & 6.1 & 7.8 \\
10.9 & 9.3 & 11.4 & 5.8 \\
9.7 & 8.8 & 6.7 & 13.2
\end{array}
\]

13. Quality control. The lives (in hours of continuous use) of 100 randomly selected flashlight batteries are

\[
\begin{array}{cc}
\text{Interval} & \text{Frequency} \\
6.95-7.45 & 2 \\
7.45-7.95 & 10 \\
7.95-8.45 & 23 \\
8.45-8.95 & 30 \\
8.95-9.45 & 21 \\
9.45-9.95 & 13 \\
9.95-10.45 & 1
\end{array}
\]

14. Stock analysis. The price–earnings ratios of 100 randomly selected stocks from the New York Stock Exchange are

\[
\begin{array}{c|c}
\text{Interval} & \text{Frequency} \\
-0.5-4.5 & 5 \\
4.5-9.5 & 54 \\
9.5-14.5 & 25 \\
14.5-19.5 & 13 \\
19.5-24.5 & 0 \\
24.5-29.5 & 1 \\
29.5-34.5 & 2
\end{array}
\]

Life Sciences

15. Medicine. The reaction times (in minutes) of a drug given to a random sample of 12 patients are

\[
\begin{array}{cccc}
4.9 & 6.4 & 5.0 \\
5.1 & 3.4 & 5.6 \\
3.9 & 5.8 & 5.8 \\
4.2 & 6.1 & 4.6
\end{array}
\]

16. Nutrition: animals. The mouse weights (in grams) of a random sample of 100 mice involved in a nutrition experiment are

\[
\begin{array}{c|c}
\text{Interval} & \text{Frequency} \\
41.5-43.5 & 3 \\
43.5-45.5 & 7 \\
45.5-47.5 & 13 \\
47.5-49.5 & 17 \\
49.5-51.5 & 19 \\
51.5-53.5 & 17 \\
53.5-55.5 & 15 \\
55.5-57.5 & 7 \\
57.5-59.5 & 2
\end{array}
\]
Section 8-4 Bernoulli Trials and Binomial Distributions

17. Reading scores. The grade-level reading scores from a reading test given to a random sample of 12 students in an urban high school graduating class are

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

18. Grade-point average. The grade-point averages of a random sample of 100 students from the graduating class of a large university are

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.95–3.15</td>
<td>6</td>
</tr>
<tr>
<td>3.15–3.35</td>
<td>5</td>
</tr>
<tr>
<td>3.35–3.55</td>
<td>4</td>
</tr>
<tr>
<td>3.55–3.75</td>
<td>3</td>
</tr>
<tr>
<td>3.75–3.95</td>
<td>2</td>
</tr>
</tbody>
</table>

In Section 8-1 we discussed frequency and relative frequency distributions, which were represented by tables and histograms (see Table 4, page 497, and Fig. 11, page 499). Frequency distributions and their corresponding probability distributions based on actual observations are empirical in nature. But there are many situations in which it is of interest (and possible) to determine the kind of relative frequency distribution we might expect before any data have actually been collected. What we have in mind is a theoretical, or hypothetical, probability distribution—that is, a probability distribution based on assumptions and theory rather than actual observations or measurements. Theoretical probability distributions are used to approximate properties of real-world distributions, assuming the theoretical and empirical distributions are closely matched.

There are many interesting theoretical probability distributions. One of particular interest because of its widespread use is the binomial distribution. The reason for the name “binomial distribution” is that the distribution is closely related to the binomial expansion of \((q + p)^n\), where \(n\) is a natural number. We start the discussion with a particular type of experiment called a Bernoulli experiment, or trial.

**Bernoulli Trials**

If we toss a coin, either a head occurs or it does not. If we roll a die, either a 3 shows or it fails to show. If you are vaccinated for smallpox, either you contract smallpox or you do not. What do all these situations have in common? All can be classified as experiments with two possible outcomes, each the complement of the other. An experiment for which there are only two possible outcomes, \(E\) or \(E'\), is called a **Bernoulli experiment**, or **trial**, named after Jacob Bernoulli (1654–1705), the Swiss scientist and mathematician who was one of the first to study systematically the probability problems related to a two-outcome experiment.
In a Bernoulli experiment or trial, it is customary to refer to one of the two outcomes as a **success** \( S \) and to the other as a **failure** \( F \). If we designate the probability of success by 

\[
P(S) = p
\]

then the probability of failure is 

\[
P(F) = 1 - p = q \quad \text{Note: } p + q = 1
\]

**Example 1**  
**Probability of Success in a Bernoulli Trial**  
Suppose that we roll a fair die and ask for the probability of a 6 turning up. This can be viewed as a Bernoulli trial by identifying a success with a 6 turning up and a failure with any of the other numbers turning up. Thus,

\[
p = \frac{1}{6} \quad \text{and} \quad q = 1 - \frac{1}{6} = \frac{5}{6}
\]

Find \( p \) and \( q \) for a single roll of a fair die, where a success is a number divisible by 3 turning up.

Now, suppose that a Bernoulli trial is repeated a number of times. It becomes of interest to try to determine the probability of a given number of successes out of the given number of trials. For example, we might be interested in the probability of obtaining exactly three 5’s in six rolls of a fair die or the probability that 8 people will not catch influenza out of the 10 who have been inoculated.

Suppose that a Bernoulli trial is repeated five times so that each trial is completely independent of any other and \( p \) is the probability of success on each trial. Then the probability of the outcome \( SSFFS \) would be 

\[
P(SSFFS) = P(S)P(S)P(F)P(F)P(S) \quad \text{See Section 7-3.}
\]

\[
= ppqqp = p^3q^2
\]

In general, we define a **sequence of Bernoulli trials** as follows:

**Definition**  
**Bernoulli Trials**  
A sequence of experiments is called a **sequence of Bernoulli trials**, or a **binomial experiment**, if

1. Only two outcomes are possible on each trial.
2. The probability of success \( p \) for each trial is a constant (probability of failure is then \( q = 1 - p \)).
3. All trials are independent.

The reason for calling a **sequence of Bernoulli trials** a **binomial experiment** will be made clear shortly.

**Example 2**  
**Probability of an Outcome of a Binomial Experiment**  
If we roll a fair die five times and identify a success in a single roll with a 1 turning up, what is the probability of the sequence \( SFFSS \) occurring?

**Solution**

\[
p = \frac{1}{6} \quad q = 1 - p = \frac{5}{6}
\]

\[
P(SFFSS) = pqqpp
\]

\[
= p^3q^2
\]

\[
= \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx .003
\]
In Example 2, find the probability of the outcome $FSSSF$. 

If we roll a fair die five times, what is the probability of obtaining exactly three 1’s? Notice how this problem differs from Example 2. In that example we looked at only one way three 1’s can occur. Then in Matched Problem 2 we saw another way. Thus, exactly three 1’s may occur in the following two sequences (among others):

$SFFSS, FSSSF$

We found that the probability of each sequence occurring is the same, namely,

$$\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$$

How many more sequences will produce exactly three 1’s? To answer this question, think of the number of ways the following five blank positions can be filled with three $S$’s and two $F$’s:

$$b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5$$

A given sequence is determined, of course, once the $S$’s are located. Thus, we are interested in the number of ways three blank positions can be selected for the $S$’s out of the five available blank positions $b_1, b_2, b_3, b_4,$ and $b_5$. This problem should sound familiar—it is just the problem of finding the number of combinations of 5 objects taken 3 at a time; that is, $C_{5,3}$. Thus, the number of different sequences of successes and failures that produce exactly three successes (exactly three 1’s) is

$$C_{5,3} = \frac{5!}{3!2!} = 10$$

Since the probability of each sequence is the same,

$$p^3 q^2 = \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$$

and there are 10 mutually exclusive sequences that produce exactly three 1’s, we have

$$P(\text{exactly three successes}) = C_{5,3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$$

$$= \frac{5!}{3!2!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$$

$$= (10) \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 \approx 0.032$$

Reasoning in essentially the same way, the following important theorem can be proved:

**Theorem 1**

**Probability of $x$ Successes in $n$ Bernoulli Trials**

The probability of exactly $x$ successes in $n$ independent repeated Bernoulli trials, with the probability of success of each trial $p$ (and of failure $q$), is

$$P(x \text{ successes}) = C_{n,x} p^x q^{n-x} \quad (1)$$

**Example 3**

**Probability of $x$ Successes in $n$ Bernoulli Trials**

If a fair die is rolled five times, what is the probability of rolling

(A) Exactly two 3’s?

(B) At least two 3’s?
Solution  (A) Use formula (1) with $n = 5, x = 2,$ and $p = \frac{1}{6}$:

$$P(x = 2) = C_{5,2} \left( \frac{1}{6} \right)^2 \left( \frac{5}{6} \right)^3$$

$$= \frac{5!}{2!3!} \left( \frac{1}{6} \right)^2 \left( \frac{5}{6} \right)^3 \approx .161$$

(B) Notice how this problem differs from part (A). Here we have

$$P(x \geq 2) = P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$$

It is actually easier to compute the probability of the complement of this event, $P(x < 2),$ and use

$$P(x \geq 2) = 1 - P(x < 2)$$

where

$$P(x < 2) = P(x = 0) + P(x = 1)$$

We now compute $P(x = 0)$ and $P(x = 1)$:

$$P(x = 0) = C_{5,0} \left( \frac{1}{6} \right)^0 \left( \frac{5}{6} \right)^5 \quad P(x = 1) = C_{5,1} \left( \frac{1}{6} \right)^1 \left( \frac{5}{6} \right)^4$$

$$= \left( \frac{5}{6} \right)^5 \approx .402 \quad = \frac{5!}{1!4!} \left( \frac{1}{6} \right)^1 \left( \frac{5}{6} \right)^4 \approx .402$$

Thus,

$$P(x < 2) = .402 + .402 = .804$$

and

$$P(x \geq 2) = 1 - .804 = .196$$

Matched Problem 3

Using the same die experiment as in Example 3, what is the probability of rolling

(A) Exactly one 3?  (B) At least one 3?

$\mathbf{\text{Binomial Formula: Brief Review}}$

Before extending Bernoulli trials to binomial distributions it is worthwhile to
review briefly the binomial formula. (A more detailed discussion of this for-
mula can be found in Appendix B-3.) To start, let us calculate directly the first
five natural number powers of $(a + b)^n$:

$$(a + b)^1 = a + b$$

$$(a + b)^2 = a^2 + 2ab + b^2$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$

$$(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$$

In general, it can be shown that a binomial expansion is given by the well-
known **binomial formula**:

\[ (a + b)^n = C_{n,0}a^n + C_{n,1}a^{n-1}b + C_{n,2}a^{n-2}b^2 + \cdots + C_{n,n}b^n \]
Section 8-4  Bernoulli Trials and Binomial Distributions  529

EXAMPLE 4  Finding Binomial Expansions  Use the binomial formula to expand $(q + p)^3$.

Solution

$(q + p)^3 = C_{3,0}q^3 + C_{3,1}q^2p + C_{3,2}qp^2 + C_{3,3}p^3$

$= q^3 + 3q^2p + 3qp^2 + p^3$

Matched Problem 4

Use the binomial formula to expand $(q + p)^4$.

Binomial Distribution

We now generalize the discussion of Bernoulli trials to binomial distributions. We start by considering a sequence of three Bernoulli trials. Let the random variable (see Section 7-5) represent the number of successes in three trials, 0, 1, 2, or 3. We are interested in the probability distribution for this random variable.

Which outcomes of an experiment consisting of a sequence of three Bernoulli trials lead to the random variable values 0, 1, 2, and 3, and what are the probabilities associated with these values? Table 1 answers these questions.

<table>
<thead>
<tr>
<th>Simple Event</th>
<th>Probability of Simple Event</th>
<th>$X_3$ x successes in 3 trials</th>
<th>$P(X_3 = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFF</td>
<td>$qqq = q^3$</td>
<td>0</td>
<td>$q^3$</td>
</tr>
<tr>
<td>FFS</td>
<td>$qpq = q^2p$</td>
<td>1</td>
<td>$3q^2p$</td>
</tr>
<tr>
<td>FSF</td>
<td>$pqq = q^2p$</td>
<td>2</td>
<td>$3qp^2$</td>
</tr>
<tr>
<td>SFF</td>
<td>$qpq = q^2p$</td>
<td>2</td>
<td>$3qp^2$</td>
</tr>
<tr>
<td>FSS</td>
<td>$qqp = qp^2$</td>
<td>2</td>
<td>$3qp^2$</td>
</tr>
<tr>
<td>SFS</td>
<td>$ppq = qp^2$</td>
<td>3</td>
<td>$p^3$</td>
</tr>
<tr>
<td>SSS</td>
<td>$ppp = p^3$</td>
<td>3</td>
<td>$p^3$</td>
</tr>
</tbody>
</table>

The terms in the last column of Table 1 are the terms in the binomial expansion of $(q + p)^3$, as we saw in Example 4. The last two columns in Table 1 provide a probability distribution for the random variable $X_3$. Note that both conditions for a probability distribution (see Section 7-5) are met:

1. $0 \leq P(X_3 = x) \leq 1$, $x \in \{0, 1, 2, 3\}$
2. $1 = 1^3 = (q + p)^3$  
   \[ = C_{3,0}q^3 + C_{3,1}q^2p + C_{3,2}qp^2 + C_{3,3}p^3 \]
   \[ = q^3 + 3q^2p + 3qp^2 + p^3 \]
   \[ = P(X_3 = 0) + P(X_3 = 1) + P(X_3 = 2) + P(X_3 = 3) \]

Reasoning in the same way for the general case, we see why the probability distribution of a random variable associated with the number of successes in a sequence of $n$ Bernoulli trials is called a binomial distribution—the probability of each number is a term in the binomial expansion of $(q + p)^n$. For this reason, a sequence of Bernoulli trials is often referred to as a binomial experiment. In terms of a formula, which we already discussed from another point of view (see Theorem 1), we have
**DEFINITION** Binomial Distribution

\[ P(X_n = x) = P(x \text{ successes in } n \text{ trials}) = C_{n,x} p^x q^{n-x} \quad x \in \{0, 1, 2, \ldots, n\} \]

where \( p \) is the probability of success and \( q \) is the probability of failure on each trial.

Informally, we will write \( P(x) \) in place of \( P(X_n = x) \).

---

**Example 5** Constructing Tables and Histograms for Binomial Distributions

Suppose a fair die is rolled three times and a success on a single roll is considered to be rolling a number divisible by 3.

(A) Write the probability function for the binomial distribution.

(B) Construct a table for this binomial distribution.

(C) Draw a histogram for this binomial distribution.

**Solution**

(A) 

\[ p = \frac{1}{3} \quad \text{Since two numbers out of six are divisible by 3} \]

\[ q = 1 - p = \frac{2}{3} \]

\[ n = 3 \]

Hence,

\[ P(x) = P(x \text{ successes in } 3 \text{ trials}) = C_{3,x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{3-x} \]

(B) 

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( C_{3,0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3 \approx 0.30 )</td>
</tr>
<tr>
<td>1</td>
<td>( C_{3,1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^2 \approx 0.44 )</td>
</tr>
<tr>
<td>2</td>
<td>( C_{3,2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^1 \approx 0.22 )</td>
</tr>
<tr>
<td>3</td>
<td>( C_{3,3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^0 \approx 0.04 )</td>
</tr>
</tbody>
</table>

(C) 

If we actually performed the binomial experiment described in Example 5 a large number of times with a fair die, we would find that we would roll no number divisible by 3 in three rolls of a die about 30% of the time, one number divisible by 3 in three rolls about 44% of the time, two numbers divisible by 3 in three rolls about 22% of the time, and three numbers divisible by 3 in three rolls only 4% of the time. Note that the sum of all the probabilities is 1, as it should be.

The graphing utility command in Figure 2A simulates 100 repetitions of the binomial experiment in Example 5. The number of successes in each trial is stored in list \( L_1 \). From Figure 2B, which shows a histogram of \( L_1 \), we note that the empirical probability of rolling one number divisible by 3 in three rolls is \( \frac{40}{100} = 40\% \), close to the theoretical probability of 44%. The empirical probabilities of 0, 2, or 3 successes also would be close to the corresponding theoretical probabilities.
Repeat Example 5, where the binomial experiment consists of two rolls of a die instead of three rolls.

Let $X$ be a random variable with probability distribution

<table>
<thead>
<tr>
<th>$X$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$\ldots$</th>
<th>$x_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_i$</td>
<td>$p_1$</td>
<td>$p_2$</td>
<td>$\ldots$</td>
<td>$p_n$</td>
</tr>
</tbody>
</table>

In Section 7-5 we defined the expected value of $X$ to be

$$E(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n$$

The expected value of $X$ is also called the mean of the random variable $X$, often denoted by $\mu$. The standard deviation of a random variable $X$ having mean $\mu$ is defined by

$$\sigma = \sqrt{(x_1 - \mu)^2 \cdot p_1 + (x_2 - \mu)^2 \cdot p_2 + \cdots + (x_n - \mu)^2 \cdot p_n}$$

If a random variable has a binomial distribution, where $n$ is the number of Bernoulli trials, $p$ is the probability of success, and $q$ the probability of failure, then the mean and standard deviation are given by the following formulas.

**RESULT** Mean and Standard Deviation (Random Variable in a Binomial Distribution)

Mean: $\mu = np$

Standard deviation: $\sigma = \sqrt{npq}$

**Insight** Let the random variable $X_3$ denote the number of successes in a sequence of three Bernoulli trials. Then $x = 0, 1, 2, \text{ or } 3$. The expected value of $X_3$ is given by

$$E(X_3) = 0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) + 3 \cdot P(3)$$

See Table 1.

$$= 0 + 1 \cdot 3q^2p + 2 \cdot 3qp^2 + 3 \cdot p^3$$

Factor out $3p$.

$$= 3p(q^2 + 2qp + p^2)$$

Factor the perfect square.

$$= 3p(q + p)^2$$

$q + p = 1$

$$= 3p$$

This proves the formula $\mu = np$ in the case $n = 3$. Similar but more complicated computations can be used to justify the general formulas $\mu = np$ and $\sigma = \sqrt{npq}$ for the mean and standard deviation of random variables having binomial distributions.
EXAMPLE 6  Computing the Mean and Standard Deviation of a Binomial Distribution  Compute the mean and standard deviation for the random variable in Example 5.

**Solution**

\[ n = 3 \quad p = \frac{1}{3} \quad q = 1 - \frac{1}{3} = \frac{2}{3} \]

\[ \mu = np = 3\left(\frac{1}{3}\right) = 1 \quad \sigma = \sqrt{npq} = \sqrt{3\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)} \approx .82 \]

Matched Problem 6  Compute the mean and standard deviation for the random variable in Matched Problem 5.

EXPLORE & DISCUSS 1

Let \( X_{100} \) denote the number of successes of 100 Bernoulli trials, each with probability of success \( p \).

(A) For what values of \( p \) would the mean of \( X_{100} \) be equal to 0? 50? 100?

(B) For what values of \( p \) would the standard deviation of \( X_{100} \) be equal to 0? 5? 10?

Application

Binomial experiments are associated with a wide variety of practical problems: industrial sampling, drug testing, genetics, epidemics, medical diagnosis, opinion polls, analysis of social phenomena, qualifying tests, and so on. Several types of applications are included in Exercise 8-4. We will now consider one application in detail.

EXAMPLE 7  Patient Recovery  The probability of recovering after a particular type of operation is .5. Let us investigate the binomial distribution involving eight patients undergoing this operation.

(A) Write the function defining this distribution.

(B) Construct a table for the distribution.

(C) Construct a histogram for the distribution.

(D) Find the mean and standard deviation for the distribution.

**Solution**

(A) Letting a recovery be a success, we have

\[ p = .5 \quad q = 1 - p = .5 \quad n = 8 \]

Hence,

\[ P(x) = P(\text{exactly } x \text{ successes in 8 trials}) = C_{8,x}(.5)^x(.5)^{8-x} = C_{8,x}(.5)^8 \]

(B)  

\[
\begin{array}{|c|c|}
\hline
x & P(x) \\
\hline
0 & C_{8,0}(.5)^8 \approx .004 \\
1 & C_{8,1}(.5)^8 \approx .031 \\
2 & C_{8,2}(.5)^8 \approx .109 \\
3 & C_{8,3}(.5)^8 \approx .219 \\
4 & C_{8,4}(.5)^8 \approx .273 \\
5 & C_{8,5}(.5)^8 \approx .109 \\
6 & C_{8,6}(.5)^8 \approx .031 \\
7 & C_{8,7}(.5)^8 \approx .004 \\
8 & C_{8,8}(.5)^8 \approx 1 \\
\hline
\end{array}
\]

(C) 

The discrepancy in the sum is due to round-off errors.
Section 8-4  Bernoulli Trials and Binomial Distributions

(D) \( \mu = np = 8(.5) = 4 \) \( \sigma = \sqrt{npq} = \sqrt{8(.5)(.5)} \approx 1.41 \)

Repeat Example 7 for four patients.

The mean of a random variable is its expected value. Use the distribution tables for the random variables of Examples 5 and 7 to compute the expected values. Do your answers agree with the results obtained using the formula \( \mu = np \)? Explain.

Answers to Matched Problems

1. \( p = \frac{1}{3}, q = \frac{2}{3} \)
2. \( p^3q^2 = (\frac{1}{3})^3(\frac{2}{3})^2 \approx .003 \)
3. (A) .402 (B) \( 1 - P(x = 0) = 1 - .402 = .598 \)
4. \( C_4,0q^4 + C_4,1q^3p + C_4,2q^2p^2 + C_4,3qp^3 + C_4,4p^4 = q^4 + 4q^3p + 6q^2p^2 + 4qp^3 + p^4 \)
5. (A) \( P(x) = P(x \text{ successes in 2 trials}) = C_{2,(\frac{1}{2})}^2(\frac{3}{2})^{2-x}, x \in \{0, 1, 2\} \)

(B)

\[
\begin{array}{cc}
0 & \frac{1}{2} \approx .44 \\
1 & \frac{1}{3} \approx .44 \\
2 & \frac{1}{6} \approx .11 \\
\end{array}
\]

(C)

Number of successes, \( x \)

\[
\begin{array}{cc}
0 & .44 \\
1 & .44 \\
2 & .11 \\
\end{array}
\]

6. \( \mu \approx .67; \sigma \approx .67 \)

7. (A) \( P(x) = P(\text{exactly } x \text{ successes in 4 trials}) = C_{4,(\frac{1}{2})}^x(\frac{1}{2})^{4-x} \)

(B)

\[
\begin{array}{cc}
0 & .06 \\
1 & .25 \\
2 & .38 \\
3 & .25 \\
4 & .06 \\
\end{array}
\]

(C)

Number of successes, \( x \)

\[
\begin{array}{cc}
0 & .06 \\
1 & .25 \\
2 & .38 \\
3 & .25 \\
4 & .06 \\
\end{array}
\]

(D) \( \mu = 2; \sigma = 1 \)

Exercise 8-4

A  Evaluate \( C_{n,x}p^xq^{n-x} \) for the values of \( n, x, \) and \( p \) given in Problems 1–6.

1. \( n = 5, x = 1, p = \frac{1}{2} \)
2. \( n = 5, x = 2, p = \frac{1}{2} \)
3. \( n = 6, x = 3, p = .4 \)
4. \( n = 6, x = 6, p = .4 \)
5. \( n = 4, x = 3, p = \frac{3}{4} \)
6. \( n = 4, x = 3, p = \frac{1}{3} \)

In Problems 7–10, a fair coin is tossed four times. What is the probability of obtaining

7. A head on the first toss and tails on each of the other tosses?
8. Exactly one head?
9. At least three tails?
10. Tails on each of the first three tosses?

11. No heads?
12. Four heads?

In Problems 13–18, construct a histogram for the binomial distribution \( P(x) = C_{n,x}p^xq^{n-x}, \) and compute the mean and standard deviation if

13. \( n = 3, p = \frac{1}{2} \)
14. \( n = 3, p = \frac{3}{4} \)
15. \( n = 4, p = \frac{1}{2} \)
16. \( n = 5, p = \frac{1}{3} \)
17. \( n = 5, p = 0 \)
18. \( n = 4, p = 1 \)
In Problems 19–24, a fair die is rolled three times. What is the probability of obtaining

19. A 6, 5, and 6, in that order?
20. A 6, 5, and 6, in any order?
21. At least two 6’s?
22. Exactly one 6?
23. No 6’s?
24. At least one 5?

25. If a baseball player has a batting average of .350, what is the probability that the player will get the following number of hits in the next four times at bat?
(A) Exactly 2 hits
(B) At least 2 hits

26. If a true–false test with 10 questions is given, what is the probability of scoring
(A) Exactly 70% just by guessing?
(B) 70% or better just by guessing?

27. A multiple-choice test consists of 10 questions, each with choices A, B, C, D, E (of which exactly one choice is correct). Which is more likely if you simply guess at each question: that all your answers are wrong, or that at least half are right? Explain.

28. If 60% of the electorate supports the mayor, what is the probability that in a random sample of 10 voters, fewer than half support her?

Construct a histogram for each of the binomial distributions in Problems 29–32. Compute the mean and standard deviation for each distribution.

29. \( P(x) = \binom{6}{x}(0.4)^x(0.6)^{6-x} \)
30. \( P(x) = \binom{6}{x}(0.6)^x(0.4)^{6-x} \)
31. \( P(x) = \binom{8}{x}(0.3)^x(0.7)^{8-x} \)
32. \( P(x) = \binom{8}{x}(0.7)^x(0.3)^{8-x} \)

In Problems 33 and 34, use a graphing utility to construct a probability distribution table.

33. A random variable represents the number of successes in 20 Bernoulli trials, each with probability of success \( p = 0.85 \).
(A) Find the mean and standard deviation of the random variable.
(B) Find the probability that the number of successes lies within 1 standard deviation of the mean.

34. A random variable represents the number of successes in 20 Bernoulli trials, each with probability of success \( p = 0.45 \).
(A) Find the mean and standard deviation of the random variable.
(B) Find the probability that the number of successes lies within 1 standard deviation of the mean.

In Problems 35 and 36, a coin is loaded so that the probability of a head occurring on a single toss is \( \frac{3}{5} \). In five tosses of the coin, what is the probability of getting

35. All heads or all tails?
36. Exactly 2 heads or exactly 2 tails?

37. Toss a coin three times or toss three coins simultaneously, and record the number of heads. Repeat the binomial experiment 100 times and compare your relative frequency distribution with the theoretical probability distribution.

38. Roll a die three times or roll three dice simultaneously, and record the number of 5’s that occur. Repeat the binomial experiment 100 times and compare your relative frequency distribution with the theoretical probability distribution.

39. Find conditions on \( p \) that guarantee the histogram for a binomial distribution is symmetrical about \( x = n/2 \). Justify your answer.

40. Consider two binomial distributions for 1,000 repeated Bernoulli trials—the first for trials with \( p = 0.15 \), and the second for trials with \( p = 0.85 \). How are the histograms for the two distributions related? Explain.

41. A random variable represents the number of heads in ten tosses of a coin.
(A) Find the mean and standard deviation of the random variable.
(B) Use a graphing utility to simulate 200 repetitions of the binomial experiment, and compare the mean and standard deviation of the numbers of heads from the simulation to the answers for part (A).

42. A random variable represents the number of 7’s or 11’s in ten rolls of a pair of dice.
(A) Find the mean and standard deviation of the random variable.
(B) Use a graphing utility to simulate 100 repetitions of the binomial experiment, and compare the mean and standard deviation of the numbers of 7’s or 11’s from the simulation to the answers for part (A).
Applications

Business & Economics

43. **Management training.** Each year a company selects a number of employees for a management training program given by a nearby university. On the average, 70% of those sent complete the program. Out of 7 people sent by the company, what is the probability that

(A) Exactly 5 complete the program?
(B) 5 or more complete the program?

44. **Employee turnover.** If the probability of a new employee in a fast-food chain still being with the company at the end of 1 year is .6, what is the probability that out of 8 newly hired people

(A) 5 will still be with the company after 1 year?
(B) 5 or more will still be with the company after 1 year?

45. **Quality control.** A manufacturing process produces, on the average, 6 defective items out of 100. To control quality, each day a sample of 10 completed items is selected at random and inspected. If the sample produces more than 2 defective items, then the whole day’s output is inspected and the manufacturing process is reviewed. What is the probability of this happening, assuming that the process is still producing 6% defective items?

46. **Guarantees.** A manufacturing process produces, on the average, 3% defective items. The company ships 10 items in each box and wishes to guarantee no more than 1 defective item per box. If this guarantee accompanies each box, what is the probability that the box will fail to satisfy the guarantee?

47. **Quality control.** A manufacturing process produces, on the average, 5 defective items out of 100. To control quality, each day a sample of 6 completed items is selected at random and inspected. If a success on a single trial (inspection of 1 item) is finding the item defective, then the inspection of each of the 6 items in the sample constitutes a binomial experiment, which has a binomial distribution.

(A) Write the function defining the distribution.
(B) Construct a table for the distribution.
(C) Draw a histogram.
(D) Compute the mean and standard deviation.

48. **Management training.** Each year a company selects 5 employees for a management training program given at a nearby university. On the average, 40% of those sent complete the course in the top 10% of their class. If we consider an employee finishing in the top 10% of the class a success in a binomial experiment, then for the 5 employees entering the program there exists a binomial distribution involving \( P(x \text{ successes out of 5}) \).

(A) Write the function defining the distribution.
(B) Construct a table for the distribution.
(C) Draw a histogram.
(D) Compute the mean and standard deviation.

Life Sciences

49. **Medical diagnosis.** A person with tuberculosis is given a chest x ray. Four tuberculosis x-ray specialists examine each x ray independently. If each specialist can detect tuberculosis 80% of the time when it is present, what is the probability that at least 1 of the specialists will detect tuberculosis in this person?

50. **Harmful side effects of drugs.** A pharmaceutical laboratory claims that a drug it produces causes serious side effects in 20 people out of 1,000, on the average. To check this claim, a hospital administers the drug to 10 randomly chosen patients and finds that 3 suffer from serious side effects. If the laboratory’s claims are correct, what is the probability of the hospital obtaining these results?

51. **Genetics.** The probability that brown-eyed parents, both with the recessive gene for blue, will have a child with brown eyes is .75. If such parents have 5 children, what is the probability that they will have

(A) All blue-eyed children?
(B) Exactly 3 children with brown eyes?
(C) At least 3 children with brown eyes?

52. **Gene mutations.** The probability of gene mutation under a given level of radiation is \( 3 \times 10^{-5} \). What is the probability of the occurrence of at least 1 gene mutation if \( 10^5 \) genes are exposed to this level of radiation?

53. **Epidemics.** If the probability of a person contracting influenza on exposure is .6, consider the binomial distribution for a family of 6 that has been exposed.

(A) Write the function defining the distribution.
(B) Construct a table for the distribution.
(C) Draw a histogram.
(D) Compute the mean and standard deviation.

54. **Side effects of drugs.** The probability that a given drug will produce a serious side effect in a person using the drug is .02. In the binomial distribution for 450 people using the drug, what are the mean and standard deviation?
Social Sciences

55. Testing. A multiple-choice test is given with 5 choices (only one is correct) for each of 10 questions. What is the probability of passing the test with a grade of 70% or better just by guessing?

56. Opinion polls. An opinion poll based on a small sample can be unrepresentative of the population. To see why, let us assume that 40% of the electorate favors a certain candidate. If a random sample of 7 is asked their preference, what is the probability that a majority will favor this candidate?

57. Testing. A multiple-choice test is given with 5 choices (only one is correct) for each of 5 questions. Answering each of the 5 questions by guessing constitutes a binomial experiment with an associated binomial distribution.
   (A) Write the function defining the distribution.
   (B) Construct a table for the distribution.
   (C) Draw a histogram.
   (D) Compute the mean and standard deviation.

58. Sociology. The probability that a marriage will end in divorce within 10 years is .4. What are the mean and standard deviation for the binomial distribution involving 1,000 marriages?

59. Sociology. If the probability is .60 that a marriage will end in divorce within 20 years after its start, what is the probability that out of 6 couples just married, in the next 20 years
   (A) None will be divorced?
   (B) All will be divorced?
   (C) Exactly 2 will be divorced?
   (D) At least 2 will be divorced?

Section 8-5 Normal Distributions

Normal Distribution
- Areas Under Normal Curves
- Approximating a Binomial Distribution with a Normal Distribution

Normal Distribution

If we take the histogram for a binomial distribution, say, the one we drew for Example 7, Section 8-4 \((n = 8, p = .5)\), and join the midpoints of the tops of the rectangles with a smooth curve, we obtain the bell-shaped curve in Figure 1.

The mathematical foundation for this type of curve was established by Abraham De Moivre (1667–1754), Pierre Laplace (1749–1827), and Carl Gauss (1777–1855). The bell-shaped curves studied by these famous mathematicians are called normal curves or normal probability distributions, and their equations are completely determined by the mean \(\mu\) and standard deviation \(\sigma\) of
the distribution. Figure 2 illustrates three normal curves with different means and standard deviations.

**Insight**

The equation for a normal curve is fairly complicated:

\[
f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}
\]

where \(\pi \approx 3.1416\) and \(e \approx 2.7183\). Given the values of \(\mu\) and \(\sigma\), however, the function is completely specified, and we could plot points or use a graphing calculator to produce its graph. Substituting \(x + h\) for \(x\) produces an equation of the same form but with a different value of \(\mu\). Therefore, in the terminology of Section 1-2, any horizontal translation of a normal curve is another normal curve.

Until now we have dealt with **discrete random variables**, that is, random variables that assume a finite or a “countably infinite” number of values (we have dealt only with the finite case). Random variables associated with normal distributions are **continuous** in nature; that is, they assume all values over an interval on a real number line. These are called **continuous random variables**. Random variables associated with people’s heights, light bulb lifetimes, or the lengths of time between breakdowns of an office copier are continuous. The following is a list of some of the important properties of normal curves (normal probability distributions of a continuous random variable):

### PROPERTIES

**Normal Curves**

1. Normal curves are bell shaped and are symmetrical with respect to a vertical line.

2. The mean is at the point where the axis of symmetry intersects the horizontal axis.

3. The shape of a normal curve is completely determined by its mean and standard deviation—a small standard deviation indicates a tight clustering about the mean and thus a tall, narrow curve; a large standard deviation indicates a large deviation from the mean and thus a broad, flat curve (see Fig. 2).
PROPERTIES Normal Curves

4. Irrespective of the shape, the area between the curve and the $x$ axis is always 1.

5. Irrespective of the shape, 68.3% of the area will lie within an interval of 1 standard deviation on either side of the mean, 95.4% within 2 standard deviations on either side, and 99.7% within 3 standard deviations on either side (see Fig. 3).

The normal probability distribution is the most important of all theoretical distributions. It is at the heart of a great deal of statistical theory, and it is also a useful tool in its own right for solving practical problems. Not only does a normal curve provide a good approximation for a binomial distribution for large $n$, but it also approximates many other relative frequency distributions. For example, normal curves often provide good approximations for the relative frequency distributions for heights and weights of people, measurements of manufactured parts, scores on IQ tests, college entrance examinations, civil service tests, and measurements of errors in laboratory experiments.

Areas Under Normal Curves

To use normal curves in practical problems, we must be able to determine areas under different parts of a normal curve. Remarkably, the area under a normal curve between a mean $\mu$ and a given number of standard deviations to the right (or left) of $\mu$ is the same, regardless of the shape of the normal curve. For example, the area under the normal curve with $\mu = 3$, $\sigma = 5$ from $\mu = 3$ to $\mu + 1.5\sigma = 10.5$ is equal to the area under the normal curve with $\mu = 15$, $\sigma = 2$ from $\mu = 15$ to $\mu + 1.5\sigma = 18$ (see Fig. 4, noting that the shaded regions have the same areas, or equivalently, the same numbers of pixels). Therefore, such areas for any normal curve can be easily determined from the areas for the standard normal curve, that is, the normal curve with mean 0 and standard deviation 1. In fact, if $z$ represents the number of standard deviations that a measurement $x$ is from a mean $\mu$, then the area under a normal curve from $\mu$ to $\mu + z\sigma$ equals the area under the standard normal curve from 0 to $z$ (see Fig. 5). Table I in Appendix C lists those areas for the standard normal curve.
**Example 1** Finding Probabilities for a Normal Distribution  
A manufacturing process produces light bulbs with life expectancies that are normally distributed with a mean of 500 hours and a standard deviation of 100 hours. What percentage of the light bulbs can be expected to last between 500 and 670 hours?

**Solution**

To answer this question, we first determine how many standard deviations 670 is from 500, the mean. This is easily done by dividing the distance between 500 and 670 by 100, the standard deviation. Thus,

\[ z = \frac{670 - 500}{100} = \frac{170}{100} = 1.70 \]

That is, 670 is 1.7 standard deviations from 500, the mean. Referring to Table I, Appendix C, we see that .4554 corresponds to \( z = 1.70 \). And since the total area under a normal curve is 1, we conclude that 45.54% of the light bulbs produced will last between 500 and 670 hours (see Fig. 6).

**Matched Problem 1**

What percentage of the light bulbs can be expected to last between 500 and 750 hours?

In general, to find how many standard deviations a measurement \( x \) is from a mean \( \mu \), first determine the distance between \( x \) and \( \mu \) and then divide by \( \sigma \):

**RESULT**

\[ z = \frac{\text{distance between } x \text{ and } \mu}{\text{standard deviation}} = \frac{x - \mu}{\sigma} \]
**Example 2** Finding Probabilities for a Normal Distribution From all light bulbs produced (see Example 1), what is the probability of a light bulb chosen at random lasting between 380 and 500 hours?

**Solution** To answer this, we first find $z$:

$$z = \frac{x - \mu}{\sigma} = \frac{380 - 500}{100} = -1.20$$

It is usually a good idea to draw a rough sketch of a normal curve and insert relevant data (see Fig. 7).

![Figure 7](image)

**FIGURE 7** Light bulb life expectancy: negative $z$

Table I in Appendix C does not include negative values for $z$, but because normal curves are symmetrical with respect to a vertical line through the mean, we simply use the absolute value (positive value) of $z$ for the table. Thus, the area corresponding to $z = -1.20$ is the same as the area corresponding to $z = 1.20$, which is .3849. And since the area under the whole normal curve is 1, we conclude that the probability of a light bulb chosen at random lasting between 380 and 500 hours is .3849.

**Matched Problem 2** What is the probability of a light bulb chosen at random lasting between 400 and 500 hours?

The first graphing utility command in Figure 8A simulates the life expectancies of 100 light bulbs by generating 100 random numbers from the normal distribution with $\mu = 500, \sigma = 100$ of Example 1. The numbers are...
stored in list L₁. Note from Figure 8A that the mean and standard deviation of L₁ are close to the mean and standard deviation of the normal distribution. From Figure 8B, which shows a histogram of L₁, we note that the empirical probability that a light bulb lasts between 380 and 500 hours is

\[
\frac{11 + 13 + 14}{100} = .37
\]

which is close to the theoretical probability of .3849 computed in Example 2.

Several important properties of a continuous random variable with normal distribution are listed in the box below.

<table>
<thead>
<tr>
<th>PROPERTIES</th>
<th>Normal Probability Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>( P(a \leq x \leq b) = \text{area under the normal curve from } a \text{ to } b )</td>
</tr>
<tr>
<td>2.</td>
<td>( P(-\infty &lt; x &lt; \infty) = 1 = \text{total area under the normal curve} )</td>
</tr>
<tr>
<td>3.</td>
<td>( P(x = c) = 0 )</td>
</tr>
</tbody>
</table>

In Example 2, what is the probability of a light bulb chosen at random having a life of exactly 621 hours? The area above 621 and below the normal curve at \( x = 621 \) is 0 (a line has no width). Thus, the probability of a light bulb chosen at random having a life of exactly 621 hours is 0. However, if the number 621 is the result of rounding a number between 620.5 and 621.5 (which is most likely the case), then the answer to the question is

\[
P(620.5 \leq x \leq 621.5) = \text{area under the normal curve from 620.5 to 621.5}
\]

The area is found using the procedures outlined in Example 2.

We have just pointed out an important distinction between a continuous random variable and a discrete random variable: For a probability distribution of a continuous random variable, the probability of \( x \) assuming a single value is always 0. On the other hand, for a probability distribution of a discrete random variable, the probability of \( x \) assuming a particular value from the set of permissible values is usually a positive number between 0 and 1. Continuous probability distributions are covered in greater depth in a course in calculus.

➤ Approximating a Binomial Distribution with a Normal Distribution

You no doubt found in some of the problems in Exercise 8-4 that when a binomial random variable assumes a large number of values (that is, when \( n \) is large), the use of the probability distribution formula

\[
P(x \text{ successes in } n \text{ trials}) = C_{n,x}p^xq^{n-x}
\]

became very tedious. It would be very helpful if there was an easily computed approximation of this distribution for large \( n \). Such a distribution is found in the form of an appropriately selected normal distribution.

To clarify ideas and relationships, let us consider an example of a normal distribution approximation of a binomial distribution with a relatively small value of \( n \). Then we will consider an example with a large value of \( n \).
**Example 3**  **Market Research**  A credit card company claims that their card is used by 40% of the people buying gasoline in a particular city. A random sample of 20 gasoline purchasers is made. If the company’s claim is correct, what is the probability that

(A) From 6 to 12 people in the sample use the card?
(B) Fewer than 4 people in the sample use the card?

**Solution**  We begin by drawing a normal curve with the same mean and standard deviation as the binomial distribution (Fig. 9). A histogram superimposed on this normal curve can be used to approximate the histogram for the binomial distribution. The mean and standard deviation of the binomial distribution are

\[
\mu = np = (20)(.4) = 8 \quad n = \text{sample size}
\]

\[
\sigma = \sqrt{npq} = \sqrt{(20)(.4)(.6)} \approx 2.19 \quad p = .4 \quad (\text{from the 40% claim})
\]

![Figure 9](chart1.png)

(A) To approximate the probability that 6 to 12 people in the sample use the credit card, we find the area under the normal curve from 5.5 to 12.5. We use 5.5 rather than 6, because the rectangle in the histogram corresponding to 6 extends from 5.5 to 6.5; and, reasoning in the same way, we use 12.5 instead of 12. To use Table I in Appendix C, we split the area into two parts: \(A_1\) to the left of the mean and \(A_2\) to the right of the mean. The

![Figure 10](chart2.png)
sketch in Figure 10 is helpful. Areas $A_1$ and $A_2$ are found as follows:

$$z_1 = \frac{x - \mu}{\sigma} = \frac{5.5 - 8}{2.19} \approx -1.14 \quad A_1 = .3729$$

$$z_2 = \frac{x - \mu}{\sigma} = \frac{12.5 - 8}{2.19} \approx 2.05 \quad A_2 = .4798$$

Total area $= A_1 + A_2 = .8527$

Thus, the approximate probability that the sample will contain between 6 and 12 users of the credit card is .85 (assuming that the firm’s claim is correct).

(B) To use the normal curve to approximate the probability that the sample contains fewer than 4 users of the credit card, we must find the area $A_1$ under the normal curve to the left of 3.5. The sketch in Figure 11 is useful. Since the total area under either half of the normal curve is .5, we first use Table I in Appendix C to find the area under the normal curve from 3.5 to the mean 8, and then subtract $A_2$ from .5:

$$z = \frac{x - \mu}{\sigma} = \frac{3.5 - 8}{2.19} \approx -2.05 \quad A_2 = .4798$$

$$A_1 = .5 - A_2 = .5 - .4798 = .0202$$

Thus, the approximate probability that the sample contains fewer than 4 users of the credit card is approximately .02 (assuming that the company’s claim is correct).

In Example 3 use the normal curve to approximate the probability that in the sample there are

(A) From 5 to 9 users of the credit card
(B) More than 10 users of the card

You no doubt are wondering how large $n$ should be before a normal distribution provides an adequate approximation for a binomial distribution. Without getting too involved, the following rule-of-thumb provides a good test:

**RESULT** Rule-of-Thumb Test

Use a normal distribution to approximate a binomial distribution only if the interval $[\mu - 3\sigma, \mu + 3\sigma]$ lies entirely in the interval from 0 to $n$. 
Note that in Example 3 the interval \([\mu - 3\sigma, \mu + 3\sigma] = [1.43, 14.57]\) lies entirely within the interval from 0 to 20; hence, the use of the normal distribution was justified.

**Example 4**

**Quality Control** A company manufactures 50,000 ballpoint pens each day. The manufacturing process produces 50 defective pens per 1,000, on the average. A random sample of 400 pens is selected from each day’s production and tested. What is the probability that the sample contains

(A) At least 14 and no more than 25 defective pens?
(B) 33 or more defective pens?

**Solution**

Is it appropriate to use a normal distribution to approximate this binomial distribution? The answer is yes, since the rule-of-thumb test passes with ease:

\[
\begin{align*}
\mu &= np = 400(0.05) = 20 \\
P &= \frac{50}{1000} = 0.05 \\
\sigma &= \sqrt{npq} = \sqrt{400(0.05)(0.95)} \approx 4.36 \\
[\mu - 3\sigma, \mu + 3\sigma] &= [6.92, 33.08]
\end{align*}
\]

This interval is well within the interval from 0 to 400.

(A) To find the approximate probability of the number of defective pens in a sample being at least 14 and not more than 25, we find the area under the normal curve from 13.5 to 25.5. To use Table I in Appendix C, we split the area into an area to the left of the mean and an area to the right of the mean, as shown in Figure 12.

![FIGURE 12](image)

\[
\begin{align*}
z_1 &= \frac{x - \mu}{\sigma} = \frac{13.5 - 20}{4.36} \approx -1.49 \\
z_2 &= \frac{x - \mu}{\sigma} = \frac{25.5 - 20}{4.36} \approx 1.26
\end{align*}
\]

Total area = \(A_1 + A_2 = 0.8281\)

Thus, the approximate probability of the number of defective pens in the sample being at least 14 and not more than 25 is 0.83.
(B) Since the total area under a normal curve from the mean on is .5, we find the area $A_1$ (see Fig. 13) from Table I in Appendix C and subtract it from .5 to obtain $A_2$.

\[ z = \frac{x - \mu}{\sigma} = \frac{32.5 - 20}{4.36} \approx 2.87 \quad A_1 = .4979 \]
\[ A_2 = .5 - A_1 = .5 - .4979 = .0021 \approx .002 \]

Thus, the approximate probability of finding 33 or more defective pens in the sample is .002. If a random sample of 400 included more than 33 defective pens, then the management would conclude that either a rare event has happened and the manufacturing process is still producing only 50 defective pens per 1,000, on the average, or something is wrong with the manufacturing process and it is producing more than 50 defective pens per 1,000, on the average. The company might very well have a policy of checking the manufacturing process whenever 33 or more defective pens are found in a sample rather than believing a rare event has happened and that the manufacturing process is still running smoothly.

Suppose in Example 4 that the manufacturing process produces 40 defective pens per 1,000, on the average. What is the approximate probability that in the sample of 400 pens there are

(A) At least 10 and no more than 20 defective pens?
(B) 27 or more defective pens?

When to Use the .5 Adjustment

If we are assuming a normal probability distribution for a continuous random variable (such as that associated with heights or weights of people), then we find $P(a \leq x \leq b)$, where $a$ and $b$ are real numbers, by finding the area under the corresponding normal curve from $a$ to $b$ (see Example 2). However, if we use a normal probability distribution to approximate a binomial probability distribution, then we find $P(a \leq x \leq b)$, where $a$ and $b$ are nonnegative integers, by finding the area under the corresponding normal curve from $a - .5$ to $b + .5$ (see Examples 3 and 4).

(A) Construct a histogram of the binomial distribution with $n = 8$ and $p = .1$.
(B) Does the binomial distribution of part (A) satisfy the rule-of-thumb test?
Chapter 8  Data Description and Probability Distributions

(C) Use a graphing utility to graph the normal distribution that has the same mean and standard deviation as the binomial distribution of part (A). How does the graph compare to the histogram?

(D) Is the normal distribution a good approximation to the binomial distribution in this case? Explain.

Answers to Matched Problems

1. 49.38%  
2. .3413  
3. (A) .70  (B) .13  
4. (A) .83  (B) .004

Exercise 8-5

A

In Problems 1–6, use Table I in Appendix C to find the area under the standard normal curve from 0 to the indicated measurement.

1. 2.00  
2. 3.30  
3. 1.24  
4. 1.08  
5. −2.75  
6. −0.92

Given a normal distribution with mean 100 and standard deviation 10, in Problems 7–12 find the number of standard deviations each measurement is from the mean. Express the answer as a positive number.

7. 115  
8. 132  
9. 90  
10. 77  
11. 124.3  
12. 83.1

Using the normal distribution described for Problems 7–12 and Table I in Appendix C, find the area under the normal curve from the mean to the indicated measurement in Problems 13–18.

13. 115  
14. 132  
15. 90  
16. 77  
17. 124.3  
18. 83.1

B

Given a normal distribution with mean 70 and standard deviation 8, find the area under the normal curve above the intervals in Problems 19–26.

19. 60–80  
20. 50–90  
21. 62–74  
22. 66–78  
23. 88 or larger  
24. 90 or larger  
25. 60 or smaller  
26. 56 or smaller

Using the normal distribution described for Problems 7–12 and Table I in Appendix C, find the area under the normal curve from the mean to the indicated measurement in Problems 19–26.

27. (A) All normal distributions have the same shape.  
(B) The area above the x axis and below the normal curve is the same for all normal distributions.
45. 175 or less
46. 188 or less

To graph Problems 47–50, use a graphing utility and refer to the normal probability distribution function with mean \( \mu \) and standard deviation \( \sigma \):

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}
\]

47. Graph equation (1) with \( \sigma = 5 \) and
   (A) \( \mu = 10 \)        (B) \( \mu = 15 \)        (C) \( \mu = 20 \)
   Graph all three in the same viewing window with
   \( X_{\text{min}} = -10, X_{\text{max}} = 40, Y_{\text{min}} = 0 \), and
   \( Y_{\text{max}} = 0.1 \).
48. Graph equation (1) with \( \sigma = 4 \) and
   (A) \( \mu = 8 \)        (B) \( \mu = 12 \)        (C) \( \mu = 16 \)
   Graph all three in the same viewing window with
   \( X_{\text{min}} = -5, X_{\text{max}} = 30, Y_{\text{min}} = 0 \), and
   \( Y_{\text{max}} = 0.1 \).
49. Graph equation (1) with \( \mu = 20 \) and
   (A) \( \sigma = 2 \)        (B) \( \sigma = 4 \)
   Graph both in the same viewing window with
   \( X_{\text{min}} = 0, X_{\text{max}} = 40, Y_{\text{min}} = 0 \), and \( Y_{\text{max}} = 0.2 \).
50. Graph equation (1) with \( \mu = 18 \) and
   (A) \( \sigma = 3 \)        (B) \( \sigma = 6 \)
   Graph both in the same viewing window with
   \( X_{\text{min}} = 0, X_{\text{max}} = 40, Y_{\text{min}} = 0 \), and \( Y_{\text{max}} = 0.2 \).
51. (A) If 120 scores are chosen from a normal distribution with mean 75 and standard deviation 8, how many scores \( x \) would be expected to satisfy \( 67 \leq x \leq 83 \)?
   (B) Use a graphing utility to generate 120 scores from the normal distribution with mean 75 and standard deviation 8. Determine the number of scores \( x \) such that \( 67 \leq x \leq 83 \), and compare your results with the answer to part (A).
52. (A) If 250 scores are chosen from a normal distribution with mean 100 and standard deviation 10, how many scores \( x \) would be expected to be greater than 110?
   (B) Use a graphing utility to generate 250 scores from the normal distribution with mean 100 and standard deviation 10. Determine the number of scores greater than 110, and compare your results with the answer to part (A).

Applications

Business & Economics

53. Sales. Salespeople for a business machine company have average annual sales of $200,000, with a standard deviation of $20,000. What percentage of the salespeople would be expected to make annual sales of $240,000 or more? Assume a normal distribution.
54. Guarantees. The average lifetime for a car battery of a certain brand is 170 weeks, with a standard deviation of 10 weeks. If the company guarantees the battery for 3 years, what percentage of the batteries sold would be expected to be returned before the end of the warranty period? Assume a normal distribution.
55. Quality control. A manufacturing process produces a critical part of average length 100 millimeters, with a standard deviation of 2 millimeters. All parts deviating by more than 5 millimeters from the mean must be rejected. What percentage of the parts must be rejected, on the average? Assume a normal distribution.

56. Quality control. An automated manufacturing process produces a component with an average width of 7.55 centimeters, with a standard deviation of 0.02 centimeter. All components deviating by more than 0.05 centimeter from the mean must be rejected. What percentage of the parts must be rejected, on the average? Assume a normal distribution.
57. Marketing claims. A company claims that 60% of the households in a given community use their product. A competitor surveys the community, using a random sample of 40 households, and finds only 15 households out of the 40 in the sample using the product. If the company’s claim is correct, what is the probability of 15 or fewer households using the product in a sample of 40? Conclusion? Approximate a binomial distribution with a normal distribution.
58. Labor relations. A union representative claims 60% of the union membership will vote in favor of a particular settlement. A random sample of 100 members
is polled, and out of these, 47 favor the settlement. What is the approximate probability of 47 or fewer in a sample of 100 favoring the settlement when 60% of all the membership favor the settlement? Conclusion? Approximate a binomial distribution with a normal distribution.

Life Sciences

59. Medicine. The average healing time of a certain type of incision is 240 hours, with standard deviation of 20 hours. What percentage of the people having this incision would heal in 8 days or less? Assume a normal distribution.

60. Agriculture. The average height of a hay crop is 38 inches, with a standard deviation of 1.5 inches. What percentage of the crop will be 40 inches or more? Assume a normal distribution.

61. Genetics. In a family with 2 children, the probability that both children are girls is approximately .25. In a random sample of 1,000 families with 2 children, what is the approximate probability that 220 or fewer will have 2 girls? Approximate a binomial distribution with a normal distribution.

Social Sciences

63. Testing. Scholastic Aptitude Tests are scaled so that the mean score is 500 and the standard deviation is 100. What percentage of the students taking this test should score 700 or more? Assume a normal distribution.

64. Politics. Candidate Harkins claims a private poll shows that she will receive 52% of the vote for governor. Her opponent, Mankey, secures the services of another pollster, who finds that 470 out of a random sample of 1,000 registered voters favor Harkins. If Harkins’s claim is correct, what is the probability that only 470 or fewer will favor her in a random sample of 1,000? Conclusion? Approximate a binomial distribution with a normal distribution.

65. Grading on a curve. An instructor grades on a curve by assuming the grades on a test are normally distributed. If the average grade is 70 and the standard deviation is 8, find the test scores for each grade interval if the instructor wishes to assign grades as follows: 10% A’s, 20% B’s, 40% C’s, 20% D’s, and 10% F’s.

66. Psychology. A test devised to measure aggressive–passive personalities was standardized on a large group of people. The scores were normally distributed, with a mean of 50 and a standard deviation of 10. If we want to designate the highest 10% as aggressive, the next 20% as moderately aggressive, the middle 40% as average, the next 20% as moderately passive, and the lowest 10% as passive, what ranges of scores will be covered by these five designations?

Chapter 8 Review

Important Terms, Symbols, and Concepts

8-1 Graphing Data

Bar graphs, broken-line graphs, and pie graphs are used to present visual interpretations or comparisons of data. Large sets of quantitative data can be organized in a frequency table, generally constructed by choosing five to twenty class intervals of equal length to cover the data range. The number of measurements that fall in a given class interval is called the class frequency, and the set of all such frequencies associated with their respective classes is called a frequency distribution. The relative frequency of a class is its frequency divided by the total number of items in the data set.

A histogram is a vertical bar graph used to represent a frequency distribution. A frequency polygon is a broken-line graph obtained by joining successive midpoints of the tops of the bars in a histogram. A cumulative frequency polygon, or ogive, is obtained by plotting the cumulative frequency over the upper boundary of the corresponding class.

8-2 Measures of Central Tendency

- Mean

Ungrouped Data: If \( x_1, x_2, \ldots, x_n \) is a set of \( n \) measurements, then the mean is given by

\[
[\text{mean}] = \frac{x_1 + x_2 + \cdots + x_n}{n}
\]

The mean is denoted by \( \bar{x} \) if the data set is a sample, and by \( \mu \) if the data set is an entire population.

Grouped Data: If a data set of \( n \) measurements is grouped into \( k \) classes, and \( x_i \) is the midpoint of the \( i \)th class interval and \( f_i \) is the \( i \)th class frequency, then the mean for the
**grouped data** is given by

\[
\text{mean} = \frac{x_1f_1 + x_2f_2 + \cdots + x_kf_k}{n}
\]

The mean for grouped data is denoted by \( \bar{x} \) if the data set is a sample, and by \( \mu \) if the data set is an entire population.

- **Median**
  - *Ungrouped Data*: Arrange the measurements in ascending or descending order. If the number of measurements is odd, the **median** is the middle measurement. If the number of measurements is even, the **median** is the mean of the two middle measurements.
  - *Grouped Data*: The **median** for grouped data with no classes of frequency zero is the number such that the histogram has the same area to the left of the median as to the right of the median.

- **Mode** The **mode** is the most frequently occurring measurement in a data set. There may be a unique mode, several modes, or, if no measurement occurs more than once, essentially no mode.

### 8-3 Measures of Dispersion

- **Range** The **range** for a set of ungrouped data is the difference between the largest and the smallest values in the data set. The **range for a frequency distribution** is the difference between the upper boundary of the highest class and the lower boundary of the lowest class.

- **Standard Deviation**
  - *Ungrouped Data*: The **sample standard deviation** \( s \) of a set of \( n \) sample measurements \( x_1, x_2, \ldots, x_n \) with mean \( \bar{x} \) is given by

\[
s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}}
\]

The square of the sample standard deviation, \( s^2 \), is called the **sample variance**.

  - *Grouped Data*: Suppose a data set of \( n \) sample measurements is grouped into \( k \) classes in a frequency table, where \( x_i \) is the midpoint, \( f_i \) is the frequency of the \( i \)th class interval, and \( \bar{x} \) is the mean of the grouped data. Then the **sample standard deviation** \( s \) for the grouped data is

\[
s = \sqrt{\frac{(x_1 - \bar{x})^2f_1 + (x_2 - \bar{x})^2f_2 + \cdots + (x_k - \bar{x})^2f_k}{n - 1}}
\]

### 8-4 Bernoulli Trials and Binomial Distributions

- **Bernoulli Trials** A sequence of experiments is called a **sequence of Bernoulli trials**, or a **binomial experiment**, if
  1. Only two outcomes are possible on each trial.
  2. The probability of success \( p \) is the same for each trial (the probability of failure is then \( q = 1 - p \)).
  3. All trials are independent.

- **Binomial Distributions** Let the random variable \( X_n \) represent the number of successes in \( n \) Bernoulli trials. The probability distribution of \( X_n \), called a **binomial distribution**, is given by

\[
P(X_n = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \ldots, n
\]

where \( p \) is the probability of success and \( q \) is the probability of failure on each trial.

The **mean** and **standard deviation** of a binomial distribution are given by the formulas \( \mu = np \) and \( \sigma = \sqrt{npq} \), respectively.

### 8-5 Normal Distributions

Normal probability distributions, or **normal curves**, are continuous curves that approximate the relative frequency distributions of measurements such as IQ scores, heights and weights of people, and errors in laboratory experiments.

- **Properties of Normal Curves**
  1. Normal curves are bell-shaped and are symmetric with respect to a vertical line.
  2. The mean \( \mu \) is at the point where the axis of symmetry intersects the horizontal axis.
  3. The shape of a normal curve is completely determined by its mean \( \mu \) and standard deviation \( \sigma \).
  4. The area between a normal curve and the horizontal axis is always 1.
  5. 68.3% of the area under a normal curve lies within 1 standard deviation of the mean, 95.4% within 2 standard deviations, and 99.7% within 3 standard deviations.

- **Areas under Normal Curves** In any normal distribution, the probability \( P(a \leq x \leq b) \) that \( x \) lies between \( a \) and \( b \) is equal to the area under the normal curve from \( a \) to \( b \). Such areas can be found from a table of areas associated with the **standard normal curve**, that is, the normal curve with mean 0 and standard deviation 1. The area under any normal curve from \( \mu \) to \( x = \mu + z\sigma \) is equal to the area under the standard normal curve from 0 to \( z \), where \( z = \frac{x - \mu}{\sigma} \).

- **Approximating Binomial Distributions with Normal Curves** To approximate a binomial distribution that is associated with a sequence of \( n \) Bernoulli trials, each having probability of success \( p \), we choose a normal distribution with mean \( \mu = np \) and standard deviation \( \sigma = \sqrt{npq} \). As a rule of thumb, the normal distribution provides a reasonable approximation if the interval \([\mu - 3\sigma, \mu + 3\sigma]\) lies entirely within the interval \([0, n]\).
Review Exercise

Work through all the problems in this chapter review and check your answers in the back of the book. Answers to all review problems are there along with section numbers in italics to indicate where each type of problem is discussed. Where weaknesses show up, review appropriate sections in the text.

A

1. Graph the following data using a bar graph and a broken-line graph.

<table>
<thead>
<tr>
<th>Year</th>
<th>Percentage of Voting Age Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>62.8</td>
</tr>
<tr>
<td>1964</td>
<td>61.9</td>
</tr>
<tr>
<td>1968</td>
<td>60.9</td>
</tr>
<tr>
<td>1972</td>
<td>55.2</td>
</tr>
<tr>
<td>1976</td>
<td>53.5</td>
</tr>
<tr>
<td>1980</td>
<td>52.8</td>
</tr>
<tr>
<td>1984</td>
<td>53.3</td>
</tr>
<tr>
<td>1988</td>
<td>50.3</td>
</tr>
<tr>
<td>1992</td>
<td>55.1</td>
</tr>
<tr>
<td>1996</td>
<td>49.0</td>
</tr>
<tr>
<td>2000</td>
<td>50.7</td>
</tr>
</tbody>
</table>

2. Graph the data in the following table using two pie graphs, one for men and one for women.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–3.5</td>
<td>1</td>
</tr>
<tr>
<td>3.5–6.5</td>
<td>5</td>
</tr>
<tr>
<td>6.5–9.5</td>
<td>7</td>
</tr>
<tr>
<td>9.5–12.5</td>
<td>2</td>
</tr>
</tbody>
</table>

3. (A) Draw a histogram for the binomial distribution

\[ P(x) = C_{6.4}(.4)^x(.6)^{6-x} \]

(B) What are the mean and standard deviation?

4. For the set of sample measurements 1, 1, 2, 2, 2, 3, 4, 4, 5, find the

(A) Mean

(B) Median

(C) Mode

(D) Standard deviation

5. If a normal distribution has a mean of 100 and a standard deviation of 10, then

(A) How many standard deviations is 118 from the mean?

(B) What is the area under the normal curve between the mean and 118?

B

6. Given the sample of 25 quiz scores listed in the table below from a class of 500 students:

<table>
<thead>
<tr>
<th>Quiz Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>14, 13, 16, 15, 17</td>
</tr>
<tr>
<td>19, 15, 14, 17, 15</td>
</tr>
<tr>
<td>15, 13, 12, 14, 14</td>
</tr>
<tr>
<td>12, 14, 13, 11, 15</td>
</tr>
<tr>
<td>16, 14, 16, 17, 14</td>
</tr>
</tbody>
</table>

(A) Construct a frequency table using a class interval of width 2 starting at 9.5.

(B) Construct a histogram.

(C) Construct a frequency polygon.

(D) Construct a cumulative frequency and relative cumulative frequency table.

(E) Construct a cumulative frequency polygon.

7. For the set of grouped sample data given in the table,

(A) Find the mean.

(B) Find the standard deviation.

(C) Find the median.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–3.5</td>
<td>1</td>
</tr>
<tr>
<td>3.5–6.5</td>
<td>5</td>
</tr>
<tr>
<td>6.5–9.5</td>
<td>7</td>
</tr>
<tr>
<td>9.5–12.5</td>
<td>2</td>
</tr>
</tbody>
</table>

8. (A) Construct a histogram for the binomial distribution

\[ P(x) = C_{6.4}(.5)^x(.5)^{6-x} \]

(B) What are the mean and standard deviation?

9. What are the mean and standard deviation for a binomial distribution with \( p = .6 \) and \( n = 1,000 \)?

In Problems 10 and 11, discuss the validity of each statement. If the statement is always true, explain why. If not, give a counterexample.

10. (A) If the data set \( x_1, x_2, \ldots, x_n \) has mean \( \bar{x} \), then the data set \( x_1 + 5, x_2 + 5, \ldots, x_n + 5 \) has mean \( \bar{x} + 5 \).

(B) If the data set \( x_1, x_2, \ldots, x_n \) has standard deviation \( s \), then the data set \( x_1 + 5, x_2 + 5, \ldots, x_n + 5 \) has standard deviation \( s + 5 \).
11. (A) If $X$ represents a binomial random variable with mean $\mu$, then $P(X \geq \mu) = .5$. 
(B) If $X$ represents a normal random variable with mean $\mu$, then $P(X \geq \mu) = .5$. 
(C) The area of a histogram of a binomial distribution is equal to the area above the $x$ axis and below a normal curve.

12. If the probability of success in a single trial of a binomial experiment with 1,000 trials is .6, what is the probability of obtaining at least 550 and no more than 650 successes in 1,000 trials? [Hint: Approximate with a normal distribution.]

13. Given a normal distribution with mean 50 and standard deviation 6, find the area under the normal curve:
   (A) Between 41 and 62
   (B) From 59 on

14. A data set is formed by recording the sums when a pair of dice is rolled 100 times. A second data set is formed by again rolling a pair of dice 100 times, but recording the product, not the sum, of the two numbers.
   (A) Which of the two data sets would you expect to have the smaller standard deviation? Explain.
   (B) To obtain evidence for your answer to part (A), use a graphing utility to simulate both experiments, and compute the standard deviations of each of the two data sets.

15. For the sample quiz scores in Problem 6 above, find the mean and standard deviation using the data:
   (A) Without grouping
   (B) Grouped, with class interval of width 2 starting at 9.5

16. A fair die is rolled five times. What is the probability of rolling:
   (A) Exactly three 6’s?
   (B) At least three 6’s?

17. Two dice are rolled three times. What is the probability of getting a sum of 7 at least once?

18. Ten students take an exam worth 100 points.
   (A) Construct a hypothetical set of exam scores for the ten students in which both the median and the mode are 30 points higher than the mean.
   (B) Could the median and the mode both be 50 points higher than the mean? Explain.

19. In the last presidential election, 39% of the registered voters in a certain city actually cast ballots.
   (A) In a random sample of 20 registered voters from that city, what is the probability that exactly 8 voted in the last presidential election?
   (B) Verify by the rule-of-thumb test that the normal distribution with mean 7.8 and standard deviation 2.18 is a good approximation of the binomial distribution with $n = 20$ and $p = .39$.
   (C) For the normal distribution of part (B), $P(x = 8) = 0$. Explain the discrepancy between this result and your answer from part (A).

20. A random variable represents the number of wins in a 12-game season for a football team that has a probability of .9 of winning any of its games.
   (A) Find the mean and standard deviation of the random variable.
   (B) Find the probability that the team wins each of its 12 games.
   (C) Use a graphing utility to simulate 100 repetitions of the binomial experiment associated with the random variable, and compare the empirical probability of a perfect season with the answer to part (B).

21. Retail sales. The daily number of bad checks received by a large department store in a random sample of 10 days out of the past year was 15, 12, 17, 5, 8, 13, 5, 16, and 4. Find the
   (A) Mean
   (B) Median
   (C) Mode
   (D) Standard deviation

Applications

Business & Economics

21. Retail sales. The daily number of bad checks received by a large department store in a random sample of 10 days out of the past year was 15, 12, 17, 5, 8, 13, 5, 16, and 4. Find the
   (A) Mean
   (B) Median
   (C) Mode
   (D) Standard deviation

22. Preference survey. Find the mean, median, and/or mode, whichever are applicable, for the following employee cafeteria service survey:

<table>
<thead>
<tr>
<th>Drink Ordered with Meal</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>435</td>
</tr>
<tr>
<td>Tea</td>
<td>137</td>
</tr>
<tr>
<td>Milk</td>
<td>298</td>
</tr>
<tr>
<td>Soft drink</td>
<td>522</td>
</tr>
<tr>
<td>Milk shake</td>
<td>392</td>
</tr>
</tbody>
</table>
23. **Plant safety.** The weekly record of reported accidents in a large auto assembly plant in a random sample of 35 weeks from the past 10 years is listed below:

\[
\begin{array}{cccccccc}
34 & 33 & 36 & 35 & 37 & 31 & 37 \\
39 & 34 & 35 & 37 & 35 & 32 & 35 \\
33 & 35 & 32 & 34 & 32 & 32 & 39 \\
34 & 31 & 35 & 33 & 31 & 38 & 34 \\
36 & 34 & 37 & 34 & 36 & 39 & 34 \\
\end{array}
\]

(A) Construct a frequency and relative frequency table using class intervals of width 2 and starting at 29.5.

(B) Construct a histogram and frequency polygon.

(C) Find the mean and standard deviation for the grouped data.

24. **Personnel screening.** The scores on a screening test for new technicians are normally distributed with mean 100 and standard deviation 10. Find the approximate percentage of applicants taking the test who score

(A) Between 92 and 108

(B) 115 or higher

25. **Market research.** A newspaper publisher claims that 70% of the people in a community read their newspaper. Doubting the assertion, a competitor randomly surveys 200 people in the community. Based on the publisher’s claim (and assuming a binomial distribution):

(A) Compute the mean and standard deviation of the binomial distribution.

(B) Determine whether the rule-of-thumb test warrants the use of a normal distribution to approximate this binomial distribution.

(C) Calculate the approximate probability of finding at least 130 and no more than 155 readers in the sample.

(D) Determine the approximate probability of finding 125 or fewer readers in the sample.

(E) Use a graphing utility to graph the relevant normal distribution.

26. **Health care.** A small town has three doctors on call for emergency service. The probability that any one doctor will be available when called is .90. What is the probability that at least one doctor will be available for an emergency call?

---

**Group Activity 1**  
**Analysis of Data on Student Lifestyle**

(A) Select several quantitative variables related to student life or the economic impact of students on the surrounding community—for example, the number of hours spent studying outside of class per week, the number of long-distance telephone calls made per month, the number of ounces of alcoholic beverages consumed per week, the number of dollars spent on recreational activities per week, and so on. Interview a total of approximately 40 students to obtain data on each of the variables you have selected. Discuss how the sample of students should be selected in order to obtain an approximately random sample. Discuss how the interviews should be conducted in order to obtain reliable information.

(B) Compute the mean, median, and standard deviation for the data set corresponding to each of your quantitative variables. Compute the proportion of the sample that lies within 1, 2, and 3 standard deviations of the mean. Which of your data sets most closely approximates a normal distribution?

(C) Use histograms and/or tables and other graphs to present the results of your study to those outside your group.

**Group Activity 2**  
**Survival Rates for a Heart Transplant**

In recent years approximately 2,400 heart transplant operations have been performed annually in the United States. The American Heart Association reported that the 1-year survival rate for a heart transplant is 82.4%, the 2-year survival rate is 78.2%, and the 3-year rate is 74.6%.
Ten patients are currently awaiting a heart transplant at St. Luke’s Hospital. Assume that each of the ten undergoes the transplant surgery.

(A) Construct probability distribution tables for the random variables $X_1$, $X_2$, and $X_3$, where $X_k$ represents the number from among the ten patients who survive for at least $k$ years. (Assume that $X_k$ is binomial.)

(B) What is the probability that eight or more of the ten heart transplant recipients will survive at least 1 year? 2 years? 3 years?

(C) Use a graphing utility to simulate 100 repetitions of the binomial experiments associated with $X_1$, $X_2$, and $X_3$, and compare the results of the simulations with the answers to part (B).

(D) If 2,500 heart transplants are performed in the United States this year, what is the probability that at least 2,000 of the recipients will survive at least 1 year? 2 years? 3 years? (Approximate the appropriate binomial distributions with normal distributions.)
Chapter 8

Exercise 8-1

1. (A) and (B)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–2.5</td>
<td>1</td>
<td>1</td>
<td>.1</td>
</tr>
<tr>
<td>2.5–4.5</td>
<td>1</td>
<td>1</td>
<td>.1</td>
</tr>
<tr>
<td>4.5–6.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5–8.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.5–10.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(C) The frequency tables and histograms are identical, but the data set in part (B) is more spread out than that of part (A).

3. (A) Let $X_{min} = 1.5$, $X_{max} = 25.5$, change Xscl from 1 to 2, and multiply Ymax and Yscl by 2; change Xscl from 1 to 4, and multiply Ymax and Yscl by 4.

(B) The shape becomes more symmetrical and more rectangular.

5. China; China; South Africa; North America

7. Annual Railroad Carloadings in the United States


13. (A)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.5–24.5</td>
<td></td>
<td>1</td>
<td>.05</td>
</tr>
<tr>
<td>24.5–28.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.5–32.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.5–36.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.5–40.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.5–44.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. (A)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-0.5–4.5$</td>
<td>5</td>
<td>.05</td>
</tr>
<tr>
<td>4.5–9.5</td>
<td>54</td>
<td>.54</td>
</tr>
<tr>
<td>9.5–14.5</td>
<td>25</td>
<td>.25</td>
</tr>
<tr>
<td>14.5–19.5</td>
<td>13</td>
<td>.13</td>
</tr>
<tr>
<td>19.5–24.5</td>
<td>0</td>
<td>.00</td>
</tr>
<tr>
<td>24.5–29.5</td>
<td>1</td>
<td>.01</td>
</tr>
<tr>
<td>29.5–34.5</td>
<td>2</td>
<td>.02</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>
A-44  Answers

(D) | Class Interval | Frequency | Cumulative Frequency | Relative Cumulative Frequency |
--- | --- | --- | --- | --- |
| −0.5−4.5 | 5 | 5 | .05 |
| 4.5−9.5 | 54 | 59 | .59 |
| 9.5−14.5 | 25 | 84 | .84 |
| 14.5−19.5 | 13 | 97 | .97 |
| 19.5−24.5 | 0 | 97 | .97 |
| 24.5−29.5 | 1 | 98 | .98 |
| 29.5−34.5 | 2 | 100 | 1.00 |

\[ P(\text{PE ratio between 4.5 and 14.5}) = .79 \]

17. Annual World Population Growth

19. Carbohydrate  Protein  Fat

21. Calories

23. Public and Private Schooling in the U.S.

25. 23 and 33; the median age decreased in the 1950s and 1960s, but increased in the other decades.

27. (A) | Class Interval | Frequency | Relative Frequency |
--- | --- | --- | --- |
| 1.95−2.15 | 21 | .21 |
| 2.15−2.35 | 19 | .19 |
| 2.35−2.55 | 17 | .17 |
| 2.55−2.75 | 14 | .14 |
| 2.75−2.95 | 9 | .09 |
| 2.95−3.15 | 6 | .06 |
| 3.15−3.35 | 5 | .05 |
| 3.35−3.55 | 4 | .04 |
| 3.55−3.75 | 3 | .03 |
| 3.75−3.95 | 2 | .02 |

\[ \sum_{i=1}^{100} \text{Relative Frequency} = 1.00 \]

27. (B) | Frequency | Relative Frequency |
--- | --- | --- |
| 1.95 | 25 |
| 2.15 | 20 |
| 2.35 | 15 |
| 2.55 | 10 |
| 2.75 | 5 |
| 2.95 | 1 |
| 3.15 | 1 |
| 3.35 | 1 |
| 3.55 | 1 |
| 3.75 | .5 |
| 3.95 | .05 |

27. (C) | Frequency | Relative Frequency |
--- | --- | --- |
| 2.05 | 25 |
| 2.25 | 20 |
| 2.45 | 15 |
| 2.65 | 10 |
| 2.85 | 5 |
| 3.05 | 1 |
| 3.25 | 1 |
| 3.45 | 1 |
| 3.65 | 1 |
| 3.85 | .5 |
| 4.05 | .05 |
Exercise 8-2
1. Mean = 3; median = 3; mode = 3
3. Modal preference is chocolate.
5. Mean = 4.4
7. The median
9. (A) Close to 3.5; close to 3.5
(B) Answer depends on results of simulation.
11. (A) 175, 175, 325, 525
(B) Let the four numbers be $u$, $v$, $w$, $x$, where $u$ and $v$ are both equal to $m_1$. Choose $w$ so that the mean of $w$ and $m_1$ is $m_2$; then choose $x$ so that the mean of $u$, $v$, $w$, and $x$ is $m_1$.
13. Mean $\approx 14.7$; median $= 11.5$; mode $= 10.1$
15. Mean $= 1,045.5$ hr; median $= 1,049.5$ hr
17. Mean $= \$1,211$; median $= \$1,228$; mode $= \$1,252$
19. Mean $= 50.5$ g; median $= 50.55$ g
21. Mean $= 1,572,000$; median $= 907,500$; no mode
23. Median $= 577$

Exercise 8-3
1. 1.15
3. (A) 70%; 100%; 100%; (B) Yes; (C)
5. 2.5
7. (A) False; (B) True
9. (A) The first data set. It is more likely that the sum is close to 7, for example, than to 2 or 12.
(B) Answer depends on results of simulation.
11. $x = 4.35$; $s = 2.45$
13. $x = 8.7$ hr; $s = 0.6$ hr
15. $x = 5.1$ min; $s = 0.9$ min
17. $x = 11.1$; $s = 2.3$

Exercise 8-4
1. $\mu = .156$
3. $\sigma = .276$
5. $\mu = .395$
7. $\mu = .943$
9. $\mu = 0.16$
11. $\mu = 0$; $\sigma = 0$
13. $\mu = .75$; $\sigma = .75$
15. $\mu = 1.333$; $\sigma = .943$
17. $\mu = 0$; $\sigma = 0$
19. $P(x) = .005$; $P(x) = .074$
21. $P(x) = .579$
23. $P(x) = .311$; (B) $P(x) = .437$
25. (A) $P(x) = .107$; (B) $P(x) = .033$
27. It is more likely that all answers are wrong (.107) than that at least half are right (.033).
29. $\mu = 2.4$; $\sigma = 1.2$
31. $\mu = 2.4$; $\sigma = 1.3$
A-46 Answers

33. (A) $\mu = 17; \sigma = 1.597$  (B) .654  35. .238
37. The theoretical probability distribution is given by $P(x) = C_{x,.5}(.5)^{3-x} = C_{x,.5}(.5)^3$.
39. $p = .5$

<table>
<thead>
<tr>
<th>Number of Heads</th>
<th>Theoretical Frequency</th>
<th>Actual Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>37.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>37.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12.5</td>
<td></td>
</tr>
</tbody>
</table>

41. (A) $\mu = 5; \sigma = 1.581$  (B) Answer depends on results of simulation.
47. (A) $P(x) = C_{x,.5}(.95)^{6-x}$  (B) $P(x) = C_{x,.5}(.5)^{6-x}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.735</td>
</tr>
<tr>
<td>1</td>
<td>.232</td>
</tr>
<tr>
<td>2</td>
<td>.031</td>
</tr>
<tr>
<td>3</td>
<td>.002</td>
</tr>
<tr>
<td>4</td>
<td>.000</td>
</tr>
<tr>
<td>5</td>
<td>.000</td>
</tr>
<tr>
<td>6</td>
<td>.000</td>
</tr>
</tbody>
</table>

49. .998  51. (A) .001  (B) .264  (C) .897
53. (A) $P(x) = C_{x,.6}(.4)^{6-x}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.004</td>
</tr>
<tr>
<td>1</td>
<td>.037</td>
</tr>
<tr>
<td>2</td>
<td>.138</td>
</tr>
<tr>
<td>3</td>
<td>.276</td>
</tr>
<tr>
<td>4</td>
<td>.311</td>
</tr>
<tr>
<td>5</td>
<td>.187</td>
</tr>
<tr>
<td>6</td>
<td>.047</td>
</tr>
</tbody>
</table>

55. .000864  57. (A) $P(x) = C_{x,.2}(.8)^{5-x}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.328</td>
</tr>
<tr>
<td>1</td>
<td>.410</td>
</tr>
<tr>
<td>2</td>
<td>.205</td>
</tr>
<tr>
<td>3</td>
<td>.051</td>
</tr>
<tr>
<td>4</td>
<td>.006</td>
</tr>
<tr>
<td>5</td>
<td>.000</td>
</tr>
</tbody>
</table>

59. (A) .0041  (B) .0467  (C) .138  (D) .959

Exercise 8-5

37. Solve the inequality $np - \sqrt{npq} \geq 0$ to obtain $n \geq 81.$  39. .89  41. .16  43. .01  45. .01
47. 49.

51. (A) Approx. .82  (B) Answer depends on results of simulation.
53. 2.28%  55. 1.24%
57. .0031; either a rare event has happened or the company’s claim is false.  59. 0.82%  61. .0158  63. 2.28%
65. A’s, 80.2 or greater; B’s, 74.2–80.2; C’s, 65.8–74.2; D’s, 59.8–65.8; F’s, 59.8 or lower
Chapter 8 Review Exercise

1. (8-1)

2. (8-1)

3. (A) (B) 2.5 (C) 2 (D) 1.34 (8-2, 8-3)

4. (A) $\bar{x} = 2.7$ (B) 2.5 (C) 2 (D) $s = 1.34$ (8-2, 8-3)

5. (A) 1.8 (B) .4641 (8-5)

6. (A)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.5–11.5</td>
<td>1</td>
<td>.04</td>
</tr>
<tr>
<td>11.5–13.5</td>
<td>5</td>
<td>.20</td>
</tr>
<tr>
<td>13.5–15.5</td>
<td>12</td>
<td>.48</td>
</tr>
<tr>
<td>15.5–17.5</td>
<td>6</td>
<td>.24</td>
</tr>
<tr>
<td>17.5–19.5</td>
<td>1</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(B) Frequency Relative frequency

7. (A) $\bar{x} = 7$ (B) $s = 2.45$ (C) 7.14 (8-2, 8-3)

8. (A)

<table>
<thead>
<tr>
<th>P( )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
</tr>
<tr>
<td>2.14</td>
</tr>
<tr>
<td>2.14</td>
</tr>
<tr>
<td>0.01</td>
</tr>
</tbody>
</table>

9. $\mu = 3; \sigma = 1.22$ (8-4)

10. (A) True (B) False (8-2, 8-3)

11. (A) False (B) True (C) True (8-4, 8-5)

12. $.999$ (8-5)

13. (A) .9104 (B) .0668 (8-5)

14. (A) The first data set. Sums range from 2 to 12, but products range from 1 to 36. (B) Answers depend on results of simulation. (8-3)

15. (A) $\bar{x} = 14.6; s = 1.83$ (B) $\bar{x} = 14.6; s = 1.78$ (8-2, 8-3)

16. (A) .0322 (B) .0355 (8-4)

17. (.421 (8-4)

18. (A) 10, 10, 20, 20, 90, 90, 90, 90, 90, 90 (B) No (8-2)

19. (A) .179

(C) The normal distribution is continuous, not discrete, so the correct analogue of part (A) is $P(7.5 \leq x \leq 8.5) \approx .18$ (using Table I in Appendix C). (8-5)

20. (A) $\mu = 10.8; \sigma = 1.039$ (B) .282 (C) Answer depends on results of simulation. (8-4)

21. (A) $\bar{x} = 10$ (B) 10 (C) 5 (D) $s = 5.14$ (8-2, 8-3)

22. Modal preference is soft drink (8-2)
23. (A)  

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.5–31.5</td>
<td>3</td>
<td>.086</td>
</tr>
<tr>
<td>31.5–33.5</td>
<td>7</td>
<td>.2</td>
</tr>
<tr>
<td>33.5–35.5</td>
<td>14</td>
<td>.4</td>
</tr>
<tr>
<td>35.5–37.5</td>
<td>7</td>
<td>.2</td>
</tr>
<tr>
<td>37.5–39.5</td>
<td>4</td>
<td>.114</td>
</tr>
<tr>
<td>39.5–41.5</td>
<td>5</td>
<td>.147</td>
</tr>
</tbody>
</table>

(B) \[ x = 34.61; \sigma = 2.22 \]

(C)  

24. (A) 57.62% (B) 6.68% (8-5)  

25. (A) \( \mu = 140; \sigma = 6.48 \) (B) Yes (C) .939 (D) .0125 (E)  

26. 999 (8-4)
DATA DESCRIPTION AND PROBABILITY DISTRIBUTIONS
TO ACCOMPANY
COLLEGE MATHEMATICS
FOR BUSINESS, ECONOMICS, LIFE SCIENCES,
AND SOCIAL SCIENCES
TENTH EDITION

RAYMOND A. BARNETT
MICHAEL R. ZIEGLER
KARL E. BYLEEN