

# CHAPTER 3

---



## Visualizing Shapes of Distributions

### CONCEPTS

- Random Variable, Discrete Population, Continuous Population, Probability Distribution, Cumulative Distribution, Parameters, Shape Measures

### OBJECTIVES

- Understand how a distribution is described by its parameters
- Be able to visualize common shape measures (mean, median, standard deviation, quartiles, skewness, kurtosis)
- Understand the relationship between a probability distribution and its cumulative
- Recognize discrete and continuous random variables and their probability distributions

## Overview of Concepts

The value of a **random variable** is determined by chance. For example, your commuting time to work last Tuesday is a random variable. The universe of all possible values that the random variable can take is called the population. If a population consists only of countable values it is a **discrete population** (e.g., number of children in a family), whereas if it consists of a range of uncountable values it is a **continuous population** (e.g., the fraction of movie patrons who buy popcorn). Most discrete populations have a finite number of values (e.g., the numbers 2 to 12 representing the sum of two dice). However, as the number of values becomes very large the two types of populations become similar (e.g., all integers between 1 and 10,000).

The relative frequency with which the population values occur is given by its **probability distribution**. If the population is continuous, this distribution is an area graph. If the population is discrete, this distribution is a stick (or bar) graph. In a continuous population the area under the probability distribution between any two values gives the relative frequency, or probability, that the random variable will be between those values. In a discrete population the height of any bar (or stick) tells the relative frequency, or probability, that the random variable will equal a particular value in the population. A **cumulative distribution** gives the relative frequency that the random variable will be less than or equal to a specific value  $X_0$ . If  $X_0$  is the maximum value in the population, the cumulative distribution will equal one.

Some probability distributions are so common they have been named. Three common distributions you may have heard of are the uniform distribution (a rectangle), the triangular distribution, and the normal distribution (the shape of a bell). Many named distributions are unbounded (i.e., they do not have an upper limit and/or lower limit). The distribution and range of values the random variable can take on depends on the distribution's **parameters**. For example, a three-digit lottery number or the number on a die that is rolled both follow a uniform distribution, which has two parameters: the minimum value and the maximum value. In choosing the lottery number, these parameter values are 000 and 999. In rolling a single die, the parameter values are 1 and 6. Although the uniform distribution has two parameters, other known distributions have as few as one parameter or as many as five parameters. In this module you will be able to manipulate a computer-created probability distribution using five parameters: centrality, dispersion (number of discrete points if the population is discrete), asymmetry, peakedness, and number of modes. This created distribution enables you to visualize over 1,200 different-shaped probability distributions. All of these distributions have an upper and lower bound.

You learned in Chapter 1 that the shape of data can be characterized by a variety of descriptive statistics, such as the sample mean, mode, standard deviation, and range. Similarly, the shape of probability distributions can also be characterized by a variety of **shape measures** (since these measures are not based on sample data but on the probability distribution itself, they are not called statistics). The shape measures presented in this module are the population mean, median, mode, midrange, standard deviation, interquartile range, range, skewness, and kurtosis. Each of these terms is defined in the glossary and is reviewed in the Learning Exercises.

## Illustration of Concepts

Consider the distribution of the **random variable** “sum of numbers on two dice.” This random variable is a **discrete population** since it consists of the set of numbers 2 through 12. The **probability distribution** is shown in Figure 1. It is a triangular distribution. Its **parameters** are 2 (minimum), 12 (maximum), and 7 (mode). Its **shape measures** are shown in Figure 2. It has a single mode at 7. This is also its mean, median, and midrange, as you would expect in a symmetric unimodal distribution. Skewness is 0. Its range is 10, interquartile range is 2, and standard deviation is 2.09. Kurtosis is 2.48, indicating that the distribution is platykurtic or less peaked than the bell-shaped distribution (whose kurtosis is 3.00). The ratio of the range to the standard deviation is 4.78 (for unimodal distributions, this ratio is typically less than 5 if the distribution is platykurtic). The single mode in the probability distribution is imperceptible on the **cumulative distribution** (Figure 3), as is often the case with discrete populations that have a small number of discrete points.

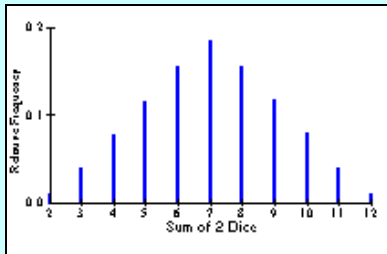


Figure 1: Discrete Population

Mean	7
Median	7
Midrange	7
Mode	7
Standard deviation	2.09
Kurtosis	2.48
Skewness	0
Range	10
Interquartile range	2
Low	2
Q1	6
Q2	7
Q3	8
High	12

Figure 2: Shape Measures

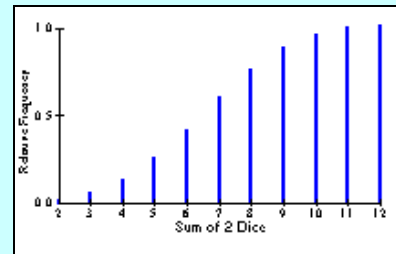


Figure 3: Cumulative Distribution

Now consider the distribution of annual incomes of females who work full time and are between the ages of 25 and 64. This is a **continuous population**, consisting of incomes from \$500 to \$95,500 (for illustrative purposes we are ignoring incomes above \$95,500 though they do exist). It is a positively skewed distribution with two modes (Figure 4). One mode is at \$14,350 and the other is at \$35,140. Since the mode at \$14,350 has the largest relative frequency, it is referred to as the global mode. It is the mode over the entire (global) population. The second mode is called a local mode since it is a mode within a portion (a localized range) of the population (about \$25,000 and above in this case). The two modes are caused by differences in education. Females with a college degree earn substantially more than those without. The population's mean is \$23,000, its median is \$18,290, and its midrange is \$48,000. The midrange is greater than the mean, which is greater than the median, as is typical of positively skewed distributions. Skewness is 1.12. The standard deviation is \$13,910, the range is \$95,000, and the interquartile range is \$19,620. Kurtosis is 4.16. The ratio of range to standard deviation is almost 7, which is consistent with kurtosis being over 3. Figure 6 shows the cumulative distribution. Both modes on the probability distribution can be seen as slight inflection points on the cumulative distribution.

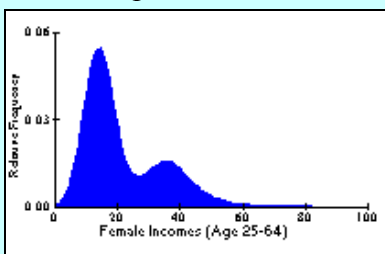


Figure 4: Continuous Population

Mean	23
Median	18.29
Midrange	48
Modes	14.35, 35.14
Standard deviation	13.91
Kurtosis	4.16
Skewness	1.12
Range	95
Interquartile range	19.62
Low	0.5
Q1	12.76
Q2	18.29
Q3	32.39
High	95.5

Figure 5: Shape Measures

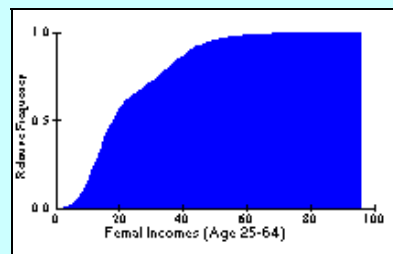


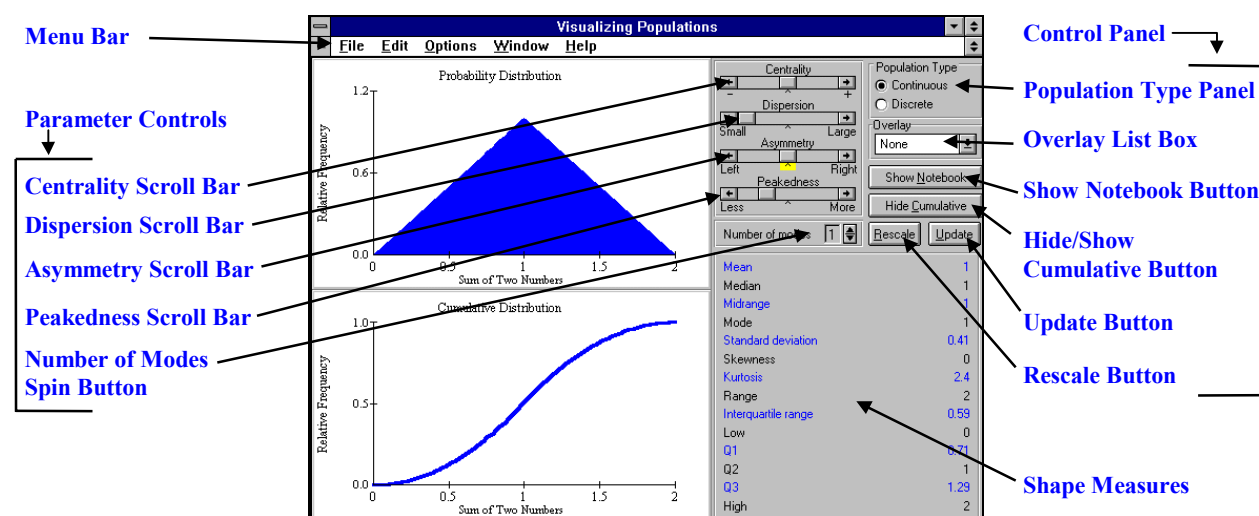
Figure 6: Cumulative Distribution

## Orientation to Basic Features

This module allows you to change the shape parameters of a computer-created distribution and see the resulting probability distribution, cumulative distribution, and shape measures. Either continuous or discrete distributions can be created. The shape measures can be visually compared.

### 1. Opening Screen

Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the **Run Module** button. When the module is loaded, you will be on the introduction page of the Notebook. Read the questions and then click the **Concepts** tab to see the concepts that you will learn. Click on the **Scenarios** tab. Select **One Mode** from the list of choices. Select a scenario, read it, and press **OK**. The upper left of the screen shows the probability distribution. Controls to change the distribution's parameters and other module options appear on the right in the Control Panel. Below the probability distribution is its cumulative distribution. The shape measures are below the Control Panel. Other features are controlled from the menu bar at the top of the screen. A flashing **Update** button indicates that you have changed one or more control settings.



### 2. Parameter Controls

The computer-created distribution's five parameters are set using either a scroll bar or a spin button located on the Control Panel. Move the **Centrality** scroll bar. Press the flashing **Update** button to redraw the displays. Try *large* changes (click on the scroll bar or drag the control button) and *small* changes (click on the arrows of the scroll bar). You can change the **Dispersion**, **Asymmetry**, and **Peakedness** parameters in the same way. The **Number of Modes** is changed using a spin button. Click on the up arrow on the spin button to increase the number of modes and the down arrow to decrease the number of modes in the probability distribution. Update the displays by pressing the **Update** button.

### 3. Rescaling

If the probability or cumulative distribution is cut off or is too small, press the **Rescale** button. This will recenter the distributions and rescale both the vertical and horizontal axes.

#### 4. **Overlays**

There are five options listed in the **Overlay** list box:

- a. No overlay (**None**) is the default.
- b. **Centrality** gives a graphic display of the population mean, median, midrange, and mode(s) overlaid on the probability distribution.
- c. **Dispersion** gives a graphic display of the population range, interquartile range, and standard deviation overlaid on the probability distribution.
- d. **Box Plot** overlays a box plot of the population values on the probability distribution. The whiskers are drawn from the smallest to the largest value in the population.
- e. **Beam & Fulcrum** overlays a beam and fulcrum of the population values on the probability distribution. The beam is drawn from the smallest to the largest value in the population.

#### 5. **Notebook**

Press the **Show Notebook** button to return to the Notebook. Click on other scenarios to read their descriptions. Any scenario can be selected by pressing **OK**. Click on the **Previous page** (upper right corner) or **Next page** (lower right corner) to page through the scenario section of the Notebook. You can return to the scenario's contents page by clicking on **Return to Scenarios contents page**. Selecting any other tab will take you to that section of the Notebook.

#### 6. **Options**

Three sets of options are available from the **Options** menu (two are discussed here, the last in the Orientation to Additional Features section) on the menu bar (top of the screen).

- a. Select **Change Title** from the **Options** menu if you wish to retitle the current display.
- b. Select **Auto Update** from the **Options** menu to automatically update each display after any change is made. Select **Auto Rescale on Update** from the **Options** menu to rescale automatically the distribution when any update occurs. Although useful at times, this option makes changes to the shape of the distribution difficult to see because the scale on the vertical and/or horizontal axis may have changed.

#### 7. **Copying a Display**

Click on the graph you wish to copy or click on the shape measures. Black handles appear indicating it has been selected. Select **Copy** from the **Edit** menu (on the menu bar at the top of the screen) or Ctrl-C to copy the display. It can then be pasted into other applications, such as Word or WordPerfect, so it can be printed. If you click on a graph and then decide not to copy it, the black handles can be removed by pressing the Esc key on your keyboard.

#### 8. **Help**

Click on **Help** on the menu bar at the top of the screen. **Search for Help** lets you search an index for this module, **Contents** shows a table of contents for this module, **Using Help** gives instructions on how to use Help, and **About** gives licensing and copyright information about *Visual Statistics*. Close Help by selecting **Exit** from the **File** menu on the Help screen.

#### 9. **Exit**

Close the module by selecting **Exit** in the **File** menu (or click  in the upper right-hand corner of the window). You will be returned to the *Visual Statistics* main menu.

## Orientation to Additional Features

### 1. Population Type

Change Population Type to **Discrete** by clicking its option button to see a discrete probability distribution. Notice that the **Dispersion** scroll bar has been renamed **Number of Discrete Points**. Clicking on the scroll bar will change the number of discrete points in the distribution (from 2 to 96). Reset Population Type to **Continuous** by clicking its option button.

### 2. Showing/Hiding the Cumulative

The cumulative distribution that corresponds to the probability distribution in the top left quadrant is automatically displayed in the lower left quadrant. Click the **Hide Cumulative** button to remove the cumulative display and enlarge the probability distribution display.

### 3. Creating Full Window Displays

Select **Full Window Graph** from the **Options** menu. This extends your distributions to the full width of your screen and hides the controls. To return to the normal display, de-select **Full Window Graph** from the **Options** menu. Use this option in conjunction with the **Hide Cumulative** button (see number 2 above) to have the probability distribution displayed on the entire screen.

## Basic Learning Exercises

Name \_\_\_\_\_

### Probability and Cumulative Distributions

1. Press the **Show Notebook** button, select the **Scenarios** tab, click on **No Mode** and select **Roll of a Die**. Read the scenario. Answer the exercises in the scenario.
2. Click **OK**. a) What does the probability distribution show? b) What does it mean in this scenario? c) What is the probability of rolling a 3?
3. a) What does a cumulative distribution show? b) What does it mean in this scenario? c) How was this cumulative distribution created? d) Calculate the cumulative probability for  $X = 5$ .
4. Press the **Show Notebook** button and click **Next page** (lower right hand corner of Notebook). Select **Sum of Two Random Numbers**, read the scenario, and answer the questions.
5. Press **OK**. How is the cumulative distribution created? What is the value at  $X = 2$ , why?
6. a) What value corresponds to the mode of the probability distribution? b) Why does the cumulative distribution equal 0.5 at this mode? c) In this scenario, what does the 0.5 mean?

**Centrality Parameter**

7. Many distributions have a centrality parameter. Use the **Centrality** scroll bar to change the value of the centrality parameter. Press the **Update** button and, if necessary, the **Rescale** button. Did changing centrality parameter change the shape of either the probability or cumulative distribution? Why or why not? This is generally true. Why?
8. Click on the **Overlay** list box and select **Centrality**. Lines representing four centrality statistics appear on the display. What is the value of the midrange, mean, median, and mode? Why are all four statistics the same? Click on the right portion of the **Asymmetry** scroll bar. Click the **Update** button. Rank these four statistics. In a unimodal, positively skewed probability distribution, this is *always* the relation between these four statistics. What is the relationship between these statistics for a negatively skewed unimodal distribution?

Midrange \_\_\_\_\_ Mean \_\_\_\_\_ Median \_\_\_\_\_ Mode \_\_\_\_\_

**Dispersion Parameter**

9. Return the **Asymmetry** scroll bar to the middle (click on “^”). a) What is the height and range of your triangular distribution? b) Click *once* on the *right arrow* of the **Dispersion** scroll bar to change the scale of the probability distribution. This is equivalent to multiplying the random variable by a number (greater than 1 if you increase dispersion, less than 1 if you decrease dispersion). Press the **Update** button. What is the height and range of your new distribution? c) Press the **Rescale** button. How did changing dispersion affect the height, range, and shape of the probability and cumulative distributions? This is generally true.
10. Press the **Show Notebook** button and press **OK** to return to the original display. Click on the **Overlay** list box and select **Dispersion**. Lines representing the three dispersion statistics appear on the display. What is the value of the standard deviation, range and interquartile range? How many standard deviations and how many interquartile ranges cover the range?

Standard Deviation \_\_\_\_\_ Interquartile Range \_\_\_\_\_ Range \_\_\_\_\_  
 Range ÷ Standard Deviation \_\_\_\_\_ Range ÷ Interquartile Range \_\_\_\_\_

## Intermediate Learning Exercises

Name \_\_\_\_\_

## Continuous Distribution

11. Press the **Show Notebook** button, select the **Scenarios** tab, click on **No Mode** and select **Random Numbers between 0 and 2**. Answer the questions. Why is its height  $\frac{1}{2}$ ?
12. Press **OK**. Describe the cumulative distribution? Calculate its value for  $X = 0.5$  and  $1.2$ .
13. a) Using the probability distribution, what is the  $P(1 \leq X \leq 2)$ ? b)  $P(1 \leq X \leq 1.1)$ ?  
c)  $P(1 \leq X \leq 1.01)$ ? d) Why does the  $P(1 \leq X \leq 1) = P(X = 1) = 0$ ? e) Why is the probability that  $X$  equals a number,  $X_0$ , always equal to 0 in a continuous distribution?

## Asymmetry Parameter

14. Press the **Show Notebook** button, select the **Scenarios** tab, click on **One Mode**, select **Sum of Two Random Numbers**, and press **OK**. Click on **Options** on the menu bar and select **Auto Update**. Create a positively skewed probability distribution by clicking on the right side of the **Asymmetry** scroll bar. Describe both distributions. Create negative skewness and describe both the probability and cumulative distribution. **Hint:** Watch the mode.

## Peakedness Parameter

15. Click on “^” on the **Peakedness** scroll bar and on the **Asymmetry** scroll bar. This produces the bell-shaped curve. What is the value of the standard deviation, range, and interquartile range? How many standard deviations and interquartile ranges cover the range?

Standard Deviation \_\_\_\_\_ Interquartile Range \_\_\_\_\_ Range \_\_\_\_\_  
 Range  $\div$  Standard Deviation \_\_\_\_\_ Range  $\div$  Interquartile Range \_\_\_\_\_

16. Set the peakedness parameter to its maximum to produce a peaked or *leptokurtic* distribution. What is the value of the standard deviation, range, and interquartile range? How many standard deviations and interquartile ranges cover the range? Describe both distributions.

Standard Deviation \_\_\_\_\_ Interquartile Range \_\_\_\_\_ Range \_\_\_\_\_  
 Range ÷ Standard Deviation \_\_\_\_\_ Range ÷ Interquartile Range \_\_\_\_\_

17. Set the peakedness parameter to its minimum to produce a flat or *platykurtic* distribution. What is the value of the standard deviation, range, and interquartile range? How many standard deviations and interquartile ranges cover the range? Describe both distributions.

Standard Deviation \_\_\_\_\_ Interquartile Range \_\_\_\_\_ Range \_\_\_\_\_  
 Range ÷ Standard Deviation \_\_\_\_\_ Range ÷ Interquartile Range \_\_\_\_\_

18. Note that the number of standard deviations and interquartile ranges needed to cover the range increases and decreases as peakedness increases and decreases. This is generally true. Why?

### Shape Measures or Shape Statistics

19. Click on “^” on all four scroll bars. Record the shape statistics. Change the **Centrality** scroll bar. Which statistic changes? Return the mean to 1. Change the **Dispersion** scroll bar. Which statistic changes? **Hint:** Recall from exercises 7 and 9 that changing centrality or dispersion does not change the shape of the distribution only its location and spread. The bell-shaped curve always has a skewness of 0 and a kurtosis of 3.0 (called *mesokurtic*).

Mean \_\_\_\_\_ Standard Deviation \_\_\_\_\_ Skewness \_\_\_\_\_ Kurtosis \_\_\_\_\_

20. Return all scroll bars to “^”. Change the **Asymmetry** scroll bar. Why do all 4 statistics change when the symmetry changes? Return **Asymmetry** to “^”. Change the **Peakedness** scroll bar. Why does the standard deviation change when the peakedness changes?

## Advanced Learning Exercises

Name \_\_\_\_\_

### Probability Distributions with Zero, Two and Three Modes

21. Press the **Show Notebook** button, select the **Scenarios** tab, click on **No Mode**, and select **Random Numbers Between 0 and 2**. Read the scenario. Answer the questions. Click **OK**. Note that the **Asymmetry** and **Peakedness** scroll bars are inoperative. Why is this? The **Overlay** feature can be turned off or set to whatever option you prefer.
  
22. Press the **Show Notebook** button. Click on **Next page** to turn a Notebook page. Turn another page to see examples of scenarios with two modes. Select the **Traffic Flow** scenario, read it, and answer the questions. Click **OK**. Describe the cumulative distribution. How can you tell by looking at the cumulative distribution that the probability distribution has two modes?
  
23. Change the skewness of the probability distribution by changing the **Asymmetry** scroll bar. Notice that the modes are no longer the same height. The taller one is the global mode (maximum over the entire distribution) and the shorter one the local mode (maximum within a region). How does the shape of the probability and cumulative distribution change as the probability distribution becomes more skewed?
  
24. Return skewness to 0 by changing the **Asymmetry** scroll bar. Increase the peakedness by clicking on the right end of its scroll bar. Watch both the probability and cumulative distribution and describe how both distributions change. Note the location of the two modes.

25. Decrease peakedness by clicking on the left end of the **Peakedness** scroll bar. Watch both the probability and cumulative distribution and describe how both distributions change. Note the location of the two modes.
26. Press the **Show Notebook** button, click on **Next page** to see scenarios with three modes and select **Eating Habits**. Read the scenario and answer the questions. Click **OK**.
27. Read the other scenarios under two and three modes. Describe your own scenario that has a probability distribution with two or three modes. What is the random variable? Is it a discrete or continuous population? Sketch what you think it would look like.

### Exploring Chebyshev's Inequality

28. Chebyshev's Theorem says that, for any distribution, at least  $8/9$  of the area will lie between the mean  $\pm 3$  standard deviations and at least  $3/4$  of the area will lie between the mean  $\pm 2$  standard deviations. You will not be able to find a distribution that violates this theorem. What characteristics of a unimodal distribution will generate a probability distribution that comes closest to these limits. **Hint:** Use the **Beam & Fulcrum** in the **Overlay** list box.
29. Given your exploration in exercise 28, what are the characteristics of a unimodal distribution that has the entire distribution contained within the smallest number of standard deviations?

## Individual Learning Projects

Write a report on one of the three topics listed below. Use the cut-and-paste facilities of the module to place the appropriate graphs in your report.

1. Changing the centrality or dispersion parameter does not alter the basic shape of a probability distribution (skewness and kurtosis do not change). However, changing the asymmetry or peakedness parameter can dramatically change the mean and standard deviation of the probability distribution. Explain and illustrate these concepts using graphs. Using the midrange and range measures, show why changing the centrality parameter is equivalent to adding or subtracting a number to the random variable. Similarly, using a continuous distribution, show why changing the dispersion parameter is equivalent to multiplying the random variable by a scale factor.
2. The box plot diagram and the beam and fulcrum diagram each give a representation of the probability distribution. Describe and illustrate how a trained statistician can tell the degree of skewness and peakedness in a unimodal probability distribution by looking at these diagrams. Use at least six different probability distributions (choose these distributions carefully so that they represent a wide variety of distributions) and *both* types of diagrams in your explanation. Which diagram do you find most informative and why?
3. Using a unimodal, continuous probability distribution, reduce the dispersion to its minimum. Notice that as the range is reduced, the ordinate (height of the distribution) is increased (actually exceeding 1). At first glance, untrained observers may think this is wrong. (a) Why might they think that a probability distribution can not have a height greater than one? (b) Illustrate (using graphs with different amounts of skewness, kurtosis, and dispersion) that this is possible and describe why it happens.

## Team Learning Projects

Select one of the three projects listed below. In each case, produce a team project that is suitable for an oral presentation. Use a large poster board(s) to display your results. Graphs should be large enough for your audience to see. Each team member should be responsible for producing some of the graphs. Ask your instructor if a written report is also expected.

1. This project is for a team of three to five members. Investigate the relationship between the cumulative and the probability distribution. Each team member should select a specific setting on the **Asymmetry** scroll bar (the range of both positive and negative skewness values must be covered by the team). By changing the peakedness of the probability distribution (from minimum to maximum peakedness), each team member must generate at least five probability and cumulative distributions. Each team member should describe how peakedness affects both the probability and cumulative distributions for a specific asymmetry setting. The team should then describe how the symmetry of the probability distribution affects the cumulative distribution.
2. This project is for a team of three to five members. Investigate how discrete distributions become more like continuous ones as the number of discrete points increases. (Specify either **Box Plot** or **Beam & Fulcrum** in the **Overlay** list box to aid your investigation.) Each member should select three settings on the **Asymmetry** scroll bar (maximum right asymmetry, moderate right asymmetry, and symmetry) and one setting on the **Peakedness** scroll bar (the range of peakedness settings must be covered by the team). For each of the three pairs of settings (1) create at least three discrete probability distributions by changing the number of points in the probability distribution, and (2) create a continuous probability distribution to show the shape of the continuous version of the probability distribution and the value of its shape parameters. The team should describe what happens as the number of discrete points gets larger and how peakedness and asymmetry affect these results.
3. This project is for a team of two to three members. Investigate the Empirical Rule. The rule says that, if data is randomly selected from a probability distribution, almost all of the data will be within 3 standard deviations of the mean and approximately 95% of the data will lie within 2 standard deviations of the mean. (Use the beam and fulcrum diagram to determine where 2 and 3 standard deviations from the mean will be). Each member should select two settings on the **Asymmetry** scroll bar (the range of either right or left asymmetry settings must be covered by the team). For each asymmetry setting generate at least five probability distributions with different amounts of peakedness (the range of peakedness values from minimum to maximum must be covered by each team member). The team should describe how peakedness and asymmetry affect the accuracy of the Empirical Rule. Why doesn't changing the dispersion or centrality parameter affect the Empirical Rule?

## Self-Evaluation Quiz

1. A cumulative distribution
  - a. has a maximum of 1.
  - b. has an indeterminate maximum.
  - c. has the same maximum as the probability distribution.
  - d. defines an area of 1 under the curve.
  - e. has none of the above characteristics.
2. Which discrete variable would *most resemble* a continuous distribution?
  - a. Number of strokes required by a seasoned pro golfer on the 18th hole.
  - b. Number of daily charity solicitation calls received at a home phone.
  - c. Number of points scored by SAT test takers (800 points possible).
  - d. Number of red cards in five-card poker hands.
  - e. Number of times per day the family dog is walked.
3. Centrality can be changed without changing dispersion.
  - a. True.
  - b. False.
4. If population A has a larger standard deviation than population B
  - a. population A will have a greater range than population B.
  - b. population A will have a smaller range than population B.
  - c. skewness will be increased.
  - d. skewness will be decreased.
  - e. we cannot say which has the greater range or skewness.
5. Dispersion is often associated with which of the following?
  - a. Mean, median, and mode.
  - b. Standard deviation and interquartile range.
  - c. Mode, quartiles, and peakedness.
  - d. Skewness and kurtosis.
  - e. None of the above.
6. If the standard deviation of a distribution is doubled, then generally we would expect that
  - a. the kurtosis will increase.
  - b. the skewness will increase.
  - c. the range will increase.
  - d. the mean will increase.
  - e. all of the above will occur.

7. Which continuous variable is *least* likely to be skewed to the right by high values?
  - a. Annual income of passengers on flights from New York to London.
  - b. Weekend gambling winnings of customers at a major casino.
  - c. Accident damage losses by renters of an auto rental company.
  - d. Cost of a plain McDonald's hamburger in various U.S. cities.
  - e. Size of itemized monthly charge account items by college students.
8. A positively skewed distribution will have a cumulative distribution that
  - a. has a lazy "S" shape.
  - b. is bowed outward.
  - c. is bowed downward.
  - d. is a straight line.
  - e. is almost vertical.
9. If a probability distribution is right-skewed
  - a. the distribution has a long left tail.
  - b. the median will exceed the mean.
  - c. the distribution has a long right tail.
  - d. the standard deviation will be small.
  - e. the mode will exceed the median.
10. If a probability distribution is platykurtic
  - a. the distribution is usually skewed.
  - b. the distribution has a "peaked" appearance.
  - c. the distribution might also be skewed.
  - d. the standard deviation usually is small.
  - e. more than one of the above is correct.
11. Which of the following *must* be a symmetric probability distribution?
  - a. Unimodal distribution.
  - b. Bimodal distribution.
  - c. Trimodal distribution.
  - d. A distribution with no mode.
  - e. None of the distributions *must* be symmetric.
12. Which is *not* true of quartiles?
  - a. They define the upper and lower 25 percent of the probability distribution.
  - b. They are shown in a beam and fulcrum display.
  - c. They are shown in a box plot.
  - d. They may be used to measure dispersion.
  - e. They can be used with the minimum and maximum values to evaluate symmetry.

## Glossary of Terms

**Asymmetry** Generally, a reference to skewness. Within this module it is one of the parameters of the computer-created probability distribution.

**Beam and fulcrum** Display that plots the position of the mean (the “fulcrum”) and the standard deviation points ( $\mu \pm 1 \sigma$ ,  $\mu \pm 2 \sigma$ ,  $\mu \pm 3 \sigma$ , etc.). Its appearance reveals skewness (the longer tail will indicate the direction of skewness) and peakedness (the larger the number of standard deviation ticks on the beam the more peaked the distribution).

**Box plot** Five-number graphical display plotting the positions of the minimum, quartiles (first, second, third), and maximum along a scale representing the distribution's domain.

**Centrality** General reference to the attempt to characterize the location of the middle or “typical” values in a distribution (mean, median, mode, etc.). Within this module it is one of the parameters of the computer-created probability distribution.

**Continuous population** A distribution whose uncountable domain of values is defined over an interval on the X-axis (e.g.,  $3 \leq X \leq 9$ ).

**Cumulative distribution** A function that maps each value of a random variable to the probability that the random variable is less than or equal to that value. The function begins at 0 and rises to 1 as you move to the right (or, less commonly from 1 to 0 as you move to the left). See **Probability distribution**.

**Discrete population** A distribution whose domain is a countable set of points (e.g.,  $X = 0, 1, 2$ ).

**Dispersion** General reference to the range and “spread” of values around the center of a distribution. Within this module it is one of the parameters of the computer-created probability distribution. See **Standard deviation**, **Range**, **Interquartile range**, and **Quartiles**.

**Distribution** See **Probability distribution**.

**Interquartile range** The distance from the first quartile to the third quartile. It is a measure of dispersion. See **Quartile**.

**Kurtosis** Measure of relative peakedness of the probability distribution. For unimodal distributions  $K = 3$  is a mesokurtic distribution (normal or bell-shaped);  $K < 3$  is a platykurtic distribution (flatter than normal, with shorter tails); and  $K > 3$  is a leptokurtic distribution (more peaked than normal, with longer tails).

**Mean** Sum (or integral) of the values of the random variable weighted by their probabilities, often denoted  $\mu$ . It may be interpreted as the fulcrum (balancing point) of the distribution along the X-axis.

**Median** The value,  $X_0$ , of the random variable such that  $P(X < X_0) = 0.50$ . If a random variable has a continuous, symmetric probability distribution, the median will equal the mean. If a random variable has a discrete, symmetric probability distribution, the median may not equal the mean because the median must be a value in the population whereas the mean is often not.

**Midrange** Average of the lowest and highest values in the distribution.

**Mode** In general, a peak on the probability distribution. If there is more than one mode (two modes is called bimodal, three modes is called trimodal), the one that is the highest is called the

global mode (the most frequent population value). Other modes are called local modes (for each, the most frequent population value within a range of population values). Within this module it is one of the parameters of the computer-created probability distribution.

**Parameters** Values that define the probability distribution.

**Peakedness** Generally a reference to kurtosis. Within this module it is one of the parameters of the computer-created probability distribution.

**Probability distribution** For a discrete distribution, each value of the domain is mapped to a probability  $P(X)$ , such that  $0 \leq P(X) \leq 1$ . The sum of the probabilities is 1. For a continuous distribution, each value of the domain is mapped to a non-negative probability. The area under the probability distribution is 1. See **Cumulative distribution**.

**Quartiles** The first quartile (denoted  $Q_1$ ) is the point along the X-axis which defines the lower 25 percent of the distribution. The second quartile (denoted  $Q_2$ ) is the median. The third quartile (denoted  $Q_3$ ) is the point that defines the upper 25 percent of the distribution.

**Random variable** A variable whose value is determined by chance. Possible values are determined by the population (either continuous or discrete) from which it is drawn.

**Range** The distance from the smallest to the largest value in the distribution's domain. It is a measure of dispersion.

**Shape measures** General reference to measures that describe the shape, location, and spread of a distribution. These measures include the population mean, median, midrange, mode(s), standard deviation, range, interquartile range, quartiles, skewness, and kurtosis.

**Skewness** Measure of relative skewness of the probability distribution. Zero indicates symmetry. Positive values show a long right tail. Negative values show a long left tail.

**Standard deviation** The square root of the variance, often denoted  $\sigma$ . The larger the standard deviation the greater the dispersion or “spread” around the mean.

**Variance** Measure of dispersion about the mean. Often denoted  $\sigma^2$ . It is the sum (or integral) of the squared difference between the values of the random variable and the mean, weighted by their probabilities.

## Solutions to Self-Evaluation Quiz

1. a Do Exercises 3–6. Read both the Overview and Illustration of Concepts.
2. c Do Exercises 1–4. Read the Overview of Concepts.
3. a Do Exercises 7, 8.
4. e Do Exercises 9, 10.
5. b Do Exercises 9, 10.
6. c Do Exercises 9, 10.
7. d Do Exercises 14, 20. Read the Illustration of Concepts.
8. b Do Exercises 14, 20.
9. c Do Exercises 14, 19–23
10. c Do Exercises 15–17, 24, 25.
11. d Do Exercises 1, 2, 14, 19–23.
12. b Do Exercises 10, 19, 20, 28.