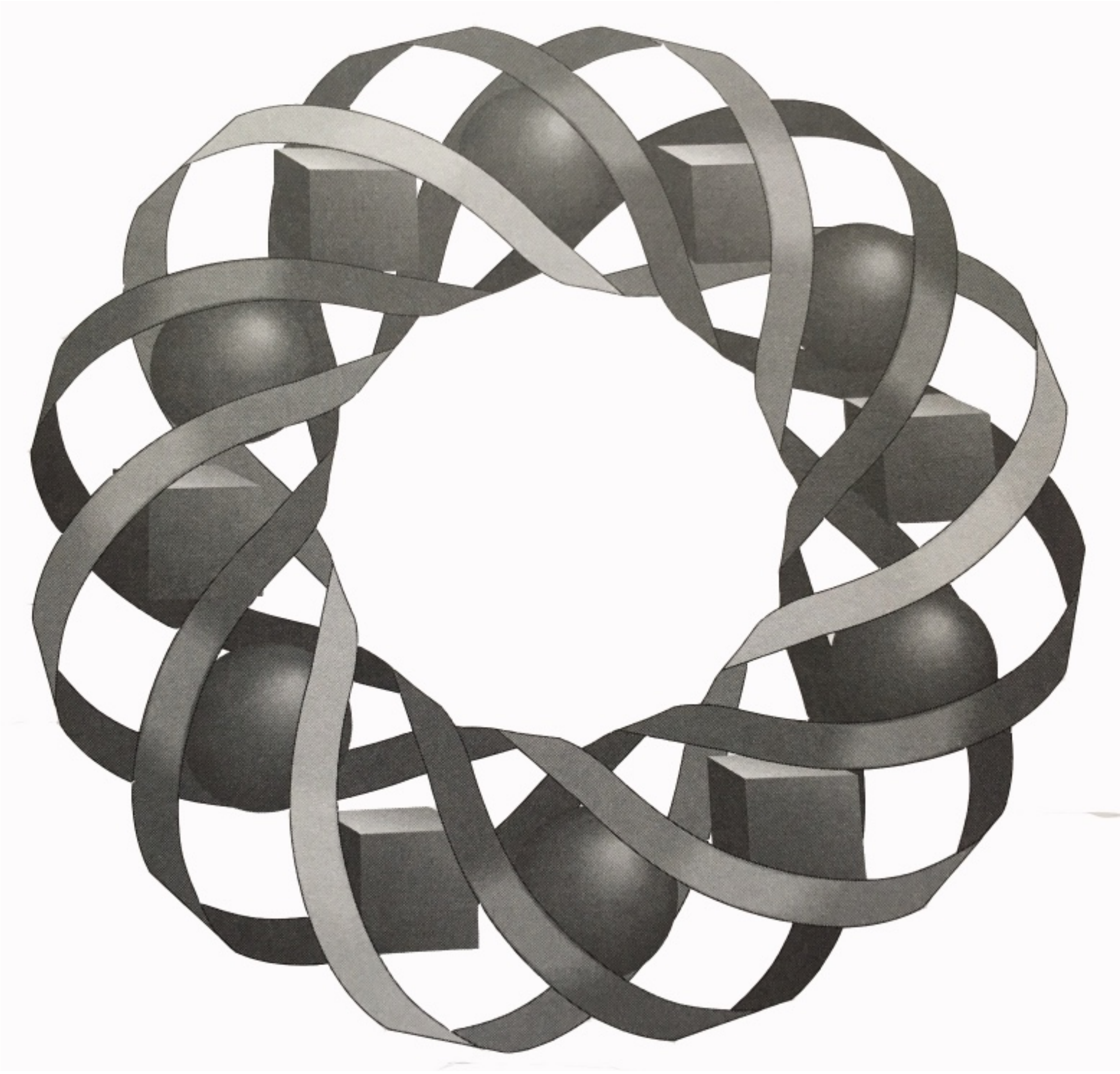
Nearly Normal



Would you recognize a normal curve if you saw one? In this module, you'll explore the mathematics of normal while examining the historic patterns of rainfall in a Northwest town.

Lee Brown • Roger Patterson • Karen Umbaugh

Nearly Normal

Introduction

The year is 2093. You have just been selected as the climate-control officer for New Bernoulli, a small city in the northwestern United States. Built near the site of the original city of Bernoulli, it is one of the first communities in America to be completely enclosed by a dome.

The dome allows for the control of New Bernoulli's air temperature, wind speed, and rainfall. Recently, the citizens have grown tired of warm, cloudless weather day after day. They miss the changing seasons. That's why they've hired you to monitor the climate.

Your job is to ensure that the following conditions are met.

- Changes in season should occur at the times of the earth's solstices and equinoxes.
- The temperature should never drop below –7˚ C and never exceed 27˚ C.
- Rain or snow showers should occur at random, although citizens must receive notice of any precipitation 24 hours in advance.
- Levels of precipitation should resemble the normal 20th-century levels.

The first three conditions seem easy to an experienced climate-control officer like yourself. The last one, however, sounds challenging. First you'll have to research the historic weather patterns for the old city of Bernoulli. Then you'll need to clarify exactly what the citizens mean by a "normal" amount of rain.

Activity 1

Your first task as climate-control officer is to determine the amount of precipitation experienced by the old city of Bernoulli in each month of a typical year. Fortunately, the National Oceanographic and Atmospheric Administration (NOAA), a federal agency that keeps records of weather patterns, had a weather station at Bernoulli throughout the 20th century.

Table **1** shows the total amount of precipitation recorded each May during the 20th century. From looking at the data, can you tell how much rain the citizens of New Bernoulli would like to receive in May of 2093?

Table 1: Recorded precipitation (in cm) for May, 1900–1999

Year	Precip.	Year	Precip.	Year	Precip.	Year	Precip.
1900	12.34	**1925**	19.67	**1950**	12.86	**1975**	16.56
1901	13.32	**1926**	11.81	**1951**	11.48	**1976**	12.44
1902	12.77	**1927**	14.53	**1952**	9.61	**1977**	7.45
1903	14.63	**1928**	9.43	**1953**	11.11	**1978**	20.49
1904	9.63	**1929**	17.43	**1954**	8.21	**1979**	5.75
1905	9.13	**1930**	11.96	**1955**	13.32	**1980**	11.07
1906	16.09	**1931**	10.68	**1956**	9.05	**1981**	16.01
1907	15.98	**1932**	9.22	**1957**	20.34	**1982**	16.11
1908	13.89	**1933**	12.41	**1958**	11.84	**1983**	10.09
1909	14.72	**1934**	18.16	**1959**	9.99	**1984**	9.49
1910	16.04	**1935**	8.59	**1960**	6.80	**1985**	13.19
1911	12.74	**1936**	11.13	**1961**	12.21	**1986**	11.34
1912	11.31	**1937**	9.19	**1962**	13.43	**1987**	12.30
1913	10.93	**1938**	14.63	**1963**	13.60	**1988**	10.22
1914	17.61	**1939**	17.78	**1964**	19.56	**1989**	13.64
1915	12.69	**1940**	13.98	**1965**	16.40	**1990**	11.19
1916	11.83	**1941**	19.85	**1966**	11.21	**1991**	18.07
1917	18.52	**1942**	21.00	**1967**	14.09	**1992**	13.39
1918	16.01	**1943**	17.86	**1968**	15.05	**1993**	11.67
1919	14.99	**1944**	11.58	**1969**	14.20	**1994**	14.56
1920	15.15	**1945**	19.85	**1970**	16.07	**1995**	16.23
1921	7.18	**1946**	13.52	**1971**	14.01	**1996**	17.89
1922	10.35	**1947**	13.65	**1972**	16.48	**1997**	19.01
1923	11.13	**1948**	13.95	**1973**	14.33	**1998**	8.89
1924	14.63	**1949**	14.02	**1974**	16.63	**1999**	8.01

Mathematics Note

The **frequency** of an item in a data set is the number of times that item is observed.

The **relative frequency** of an item is the ratio of its frequency to the total number of observations in the data set.

A **relative frequency table** includes columns that describe a data item or interval, its frequency, and its relative frequency. For example, Table **2** shows a relative frequency table for 20 observations.

Table 2: Frequency table of May precipitation (1904–1923)

Precipitation (cm)	Frequency (yr)	Relative Frequency
$[4, 6)$	0	$0/20 = 0.00$
$[6, 8)$	1	$1/20 = 0.05$
$[8, 10)$	2	$2/20 = 0.10$
$[10, 12)$	5	$5/20 = 0.25$
$[12, 14)$	3	$3/20 = 0.15$
$[14, 16)$	4	$4/20 = 0.20$
$[16, 18)$	3	$3/20 = 0.15$
$[18, 20)$	1	$1/20 = 0.05$
$[20, 22)$	1	$1/20 = 0.05$
$[22, 24)$	0	$0/20 = 0.00$

A **frequency histogram** consists of bars of equal width whose heights indicate the frequencies of intervals.

A **frequency polygon** is formed by the line graph that connects the set of points (x,y), where x is the midpoint of an interval and y is the frequency of the interval. The base of a frequency polygon is the x-axis. For example, Figure **1** shows a frequency histogram and polygon created using the data in Table **2**.

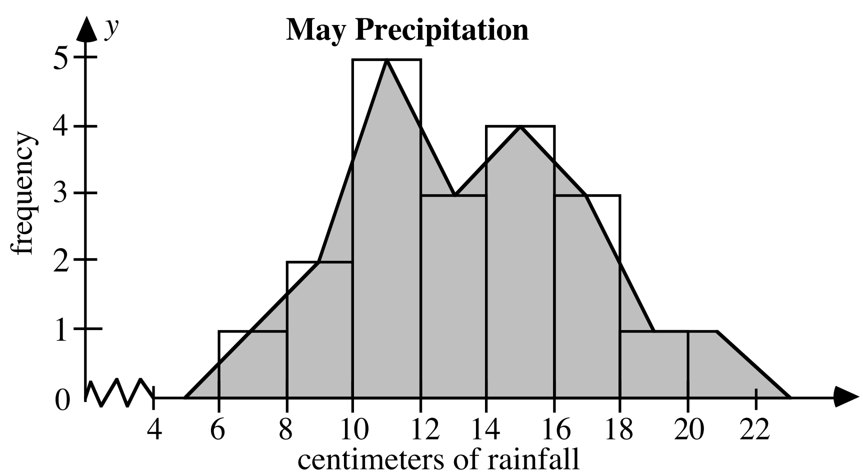


Figure 1: Frequency histogram and polygon

Exploration

In the following exploration, you organize the data in Table **1** to determine an appropriate range of rainfall amounts for New Bernoulli. **Note:** Save the data and graphs from this exploration for use later in the module.

In statistics, a **population** is the entire group of items or individuals about which information is desired. A **sample** is a portion of the population used to gather information about the whole population. Since the amounts of precipitation in years prior to 1900 are unavailable, the historic data in Table **1** may be considered a population rather than a sample.

a. Find the mean, median, and standard deviation of the data in Table **1**.

b.
1. Create a relative frequency table of Bernoulli's May precipitation from 1900 to 1999. Use intervals of 2 cm, where the first interval is $[4, 6)$.
2. Calculate the sum of the frequencies and the sum of the relative frequencies.

c.
1. Create a frequency histogram of the data.
2. Create a frequency polygon of the data.

d.
1. In a relative frequency histogram, the heights of the bars represent relative frequencies instead of frequencies. Create a relative frequency histogram of the data.
2. In a relative frequency polygon, the y-values represent relative frequencies instead of frequencies. Create a relative frequency polygon of the data.

e. On the x-axis of each graph from Parts **c** and **d**,
1. identify the x-value where the mean occurs. Label this point M.
2. identify the x-values 1 standard deviation on either side of the mean. Label the right-most value 1σ. Label left-most value -1σ
3. identify the x-values 2 standard deviations on either side of the mean. Label the right-most value 2σ. Label left-most value -2σ

Discussion

a.
1. Which interval in your table from Part **b** has the highest frequency?
2. Should this interval contain the mean and median of the data? Explain your response.

b. Weather forecasters often mention levels of precipitation that are above or below normal. What do they mean by this use of the word *normal*?

c. Describe the shapes of the frequency polygon and the relative frequency polygon of the May precipitation data.

d. Suppose that the May precipitation pattern from 1800 to 1899 was similar to that from 1900 to 1999.

1. How do you think that the shape of the frequency polygon would change if you included the May precipitation data from 1800 to 1899?

2. How do you think that the shape of the relative frequency polygon would change?

e. **1.** What is the probability that the May precipitation in New Bernoulli was in the interval [12, 14) during the 20th century? Explain your response.

2. How is the relative frequency of an interval of rainfall related to its probability?

f. What percentage of the May rainfall amounts in Table **1** are:

1. within 1 standard deviation of the mean?

2. within 2 standard deviations of the mean?

g. How could the relative frequency polygon help you determine the amount of rainfall New Bernoulli should receive in May?

Assignment

1.1 A resident of New Bernoulli has suggested randomly selecting one of the values in Table **1** for the May precipitation amount in the year 2093. Use your relative frequency table to answer the following questions.

a. What is the probability that the precipitation will be between 10 cm and 12 cm?

b. What is the probability that the precipitation will be greater than or equal to 16 cm?

c. What is the probability that the precipitation will be less than 8 cm?

1.2 What is the sum of the relative frequencies in your table? Describe how this sum is related to the probabilities associated with these relative frequencies.

1.3 Obtain a table of the annual precipitation for the past 50 years for a city near your home.

a. Calculate the mean, median, and standard deviation of the data.

b. Create a frequency table of the data, using an appropriate width for the intervals of rainfall.

c. Determine the relative frequency of each interval.

d. Create a frequency histogram of the data.

e. Create a frequency polygon of the data.

1.4 **a.** On the x-axis of your graphs from Problem **1.3**, identify and label the values for the mean, 1 standard deviation on either side of the mean, and 2 standard deviations on either side of the mean as you did in the exploration.

b. Using the data from Problem **1.3**, determine the percentage of precipitation amounts that are within:

1. 1 standard deviation of the mean
2. 2 standard deviations of the mean.

* * * * *

1.5 The following table contains the selling prices of 35 homes in New Bernoulli during the month of October.

$97,000	$126,500	$180,000	$110,000	$78,000
$134,000	$165,000	$190,000	$153,500	$118,500
$210,300	$108,000	$99,500	$85,000	$128,500
$136,000	$203,000	$116,000	$117,500	$141,900
$133,900	$217,000	$169,000	$209,900	$127,000
$181,500	$126,500	$150,000	$112,500	$119,000
$89,900	$131,400	$72,000	$137,800	$162,000

a. Calculate the mean, median, and standard deviation of the data.

b. Create a frequency table of the data, using an appropriate width for the intervals.

c. Determine the relative frequency of each interval.

d. Create a frequency histogram of the data.

e. Create a frequency polygon of the data.

1.6 **a.** On the x-axis of your graphs from Problem **1.5**, identify the x-value where the mean occurs.

b. Using the data from Problem **1.5**, determine the percentage of selling prices that are within:

1. 1 standard deviation of the mean
2. 2 standard deviations of the mean.

c. What is the probability that a house sold in October cost more than $130,000? Explain your response.

* * * * * * * * * *

Activity 2

The climate in New Bernoulli is controlled by computers. In order to recreate a rainfall pattern like that of the 20th century, you have decided to program the computers to perform a simulation.

The model should produce a set of relative frequencies that closely follows Bernoulli's historic rainfall patterns. In order to accomplish this, it may help to examine its **discrete probability distribution**.

Mathematics Note

A **probability distribution** is the assignment of probabilities to a specific characteristic that belongs to each possible outcome of an experiment. If the set of outcomes is either finite or can have a one-to-one correspondence with the natural numbers, the distribution is **discrete**.

For example, an experiment that involves tossing 3 fair coins has 8 equally likely outcomes: TTT, TTH, THT, THH, HTT, HTH, HHT, and HHH, where H represents a head and T represents a tail. One observable characteristic is the number of heads that occur. From the list of 8 possible outcomes, 1 outcome results in 0 heads, 3 outcomes result in 1 head, 3 outcomes result in 2 heads, and 1 outcome results in 3 heads.

Table **3** shows the probability distribution of the number of heads when tossing 3 fair coins (or tossing 1 fair coin 3 times). Since the number of outcomes is finite, this probability distribution is discrete.

Table 3: Probability distribution table for number of heads

No. of Heads	Frequency	Probability
0	1 (TTT)	$1/8 = 0.125$
1	3 (HTT, THT, TTH)	$3/8 = 0.375$
2	3 (HHT, HTH, THH)	$3/8 = 0.375$
3	1 (HHH)	$1/8 = 0.125$

Figure **2** shows a histogram of the probabilities associated with this experiment.

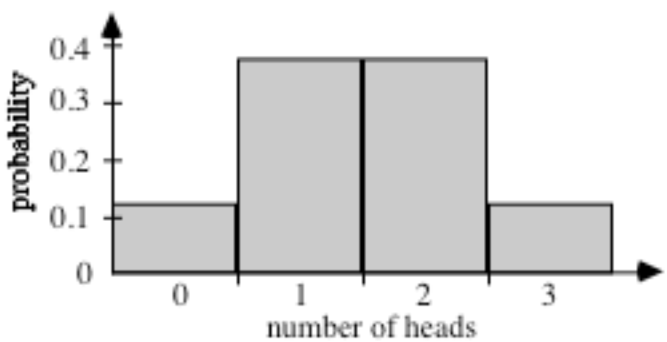


Figure 2: Probability histogram for number of heads

Exploration 1

In the following explorations, you experiment with coin tossing to create a simulation for Bernoulli's rainfall data.

a. Complete Table **4** to represent the probability distribution for the number of heads when tossing 2 fair coins.

Table 4: Probability distribution for number of heads when tossing 2 fair coins

No. of Heads	Frequency	Probability
0		
1		
2		
Total		

b. Complete Table **5** to represent the probability distribution for the number of heads when tossing 4 fair coins.

Table 5: Probability distribution for number of heads when tossing 4 fair coins

No. of Heads	Frequency	Probability
0		
1		
2		
3		
4		
Total		

c. When 2 coins are tossed, there are 4 possible outcomes. When 3 coins are tossed, there are 8 possible outcomes. When 4 coins are tossed, there are 16 possible outcomes. Use this pattern to determine the number of possible outcomes when n coins are tossed, where $n \geq 2$.

d. Complete Table **6** to show the frequencies for each number of heads when tossing 1 to 6 coins. Use the information you have already gathered in Parts **a–c** to begin the table. For example, the third row contains the frequencies when tossing 3 coins (as described in the mathematics note and in Table **3**).

Table 6: Frequencies for number of heads

No. of Coins	No. of Heads						
	0	**1**	**2**	**3**	**4**	**5**	**6**
1							
2							
3	1	3	3	1			
4							
5							
6							

e. Create a frequency histogram for each row in Table **6**. Record any similarities you observe among these histograms.

f.
1. When tossing 3 coins, the 8 possible outcomes can be listed in terms of numbers of heads. The set that represents this characteristic for the 8 possible outcomes is $\{0, 1, 1, 1, 2, 2, 2, 3\}$. Determine the mean and the standard deviation of this data set.
2. As described in Step **1**, list the possible outcomes in terms of number of heads when tossing 1, 2, 4, 5, and 6 coins.
3. Find the mean and standard deviation of each data set in Step **2**.

Discussion 1

a. Describe a pattern in Table **6** that would allow you to determine the frequencies of heads when tossing any number of coins.

b. What similarities did you observe among the histograms in Part **e** of Exploration **1**?

c. Compare the shape of the frequency histogram for tossing six coins to the frequency histogram for the rainfall data from Activity **1**.

d. Do you think that tossing one coin 6 times would produce the same probability distribution for numbers of heads as tossing six coins at once? Explain your response.

Mathematics Note

A **binomial experiment** has the following characteristics:

- It consists of a fixed number of repetitions of the same action. Each repetition is a **trial**.
- The trials are independent of each other. In other words, the result of one trial does not influence the result of any other trial in the experiment.
- Each trial has only two possible outcomes: a success or a failure.
- The probability of a success remains the same from trial to trial.
- The total number of successes is observed.

For example, consider an experiment that consists of tossing a six-sided die 10 times and observing the number of times that a 6 appears. In this case, there is a fixed number of trials, 10. For each trial, there are only two possible outcomes: either a 6 or not a 6. The probability that a 6 appears remains constant for each toss, and the result of one toss does not influence the result of any other. Therefore, this represents a binomial experiment.

The probability distribution for a binomial experiment is a **binomial distribution**. The mean of a binomial distribution is the product of the number of trials and the probability of a success. In other words, the mean μ can be found as follows, where n is the number of trials and p is the probability of a success:

$$\mu = np$$

The standard deviation σ of a binomial distribution is the square root of the product of the number of trials, the probability of a success, and the probability of a failure:

$$\sigma = \sqrt{np(1-p)}$$

For example, consider the experiment that consists of tossing a six-sided die 10 times and observing the number of times that a 6 appears. In this case, $n = 10$ and the probability of a success is $1/6$. Therefore, the mean of the corresponding binomial distribution is $10(1/6) = 10/6 \approx 1.67$. The standard deviation is $\sqrt{10(1/6)(5/6)} = \sqrt{50/36} \approx 1.18$.

e. Can an experiment that involves tossing three fair coins and counting the number of heads be considered a binomial experiment? Explain your response.

f. Consider a binomial experiment that involves tossing a fair coin six times and counting the number of heads.

1. What is n, the number of trials, in this experiment?
2. What is p the probability of a success?
3. Using the formula $\mu = np$, what is the mean of the probability distribution for this experiment?
4. How does this mean compare with the mean for tossing six coins that you calculated in Part **f** of Exploration **1**?

g. As described in the previous mathematics note, the standard deviation of a binomial distribution is the square root of the product of the number of trials, the probability of a success, and the probability of a failure:

$$\sigma = \sqrt{np(1-p)}$$

1. Why is the probability of a failure $(1 - p)$?
2. Using the formula above, what is the standard deviation for an experiment that involves tossing a fair coin six times and counting the number of heads that appear?
3. How does this value compare with the standard deviation for tossing six coins that you calculated in Part **f** of Exploration **1**?

h. Table **5** shows the theoretical probability distribution for number of heads when tossing four fair coins. If you tossed four coins 100 times, do you think that the experimental probability distribution would be the same as the theoretical one? Explain your response.

Exploration 2

The shapes of the frequency histograms for tossing coins and the shape of the frequency histogram for Bernoulli's historic rainfall data are very similar. To take advantage of this fact, you decide to use a coin tossing simulation to help determine amounts of rainfall for New Bernoulli.

a. In Activity **1**, you created a relative frequency table of the May precipitation using thirteen 2-cm intervals. To model these 13 intervals, the simulation will require 12 coins.

Place 12 coins in a container. Shake the container, pour the coins onto a table, and count the number of heads that appear.

b. In the simulation, each possible number of heads is associated with an interval of precipitation. Repeat the process described in Part **a** until you have obtained 100 observations for this experiment (to represent 100 years of data).

Record your results in Table **7** and calculate the relative frequencies for each interval.

Table 7: Simulating May precipitation using 12 coins

No. of Heads	Precipitation (cm)	Frequency	Relative Frequency
0	$[0,2)$		
1	$[2,4)$		
2	$[4,6)$		
3	$[6,8)$		
4	$[8,10)$		
5	$[10,12)$		
6	$[12,14)$		
7	$[14,16)$		
8	$[16,18)$		
9	$[18,20)$		
10	$[20,22)$		
11	$[22,24)$		
12	$[24,26)$		

c. Determine the mean, median, and standard deviation of the data in terms of number of heads.

d. Use the information in Table **7** to draw a relative frequency histogram and a relative frequency polygon.

e. On the x-axis of each graph from Part **d** above:

1. identify and label the x-value where the mean occurs
2. identify and label the x-values 1 standard deviation on either side of the mean
3. identify and label the x-values 2 standard deviations on either side of the mean.

f. Using the pattern described in Part **a** of Discussion **1**, determine the theoretical probability distribution for 12 coins. Create a relative frequency polygon of this distribution.

g. Use the formulas $\mu = np$ and $\sigma = \sqrt{np(1-p)}$ to calculate the mean and standard deviation of the binomial distribution for tossing a coin 12 times and counting the number of heads that appear. Compare these values to the ones you determined in Part **c**.

Discussion 2

a. Compare the shapes of the relative frequency polygons you created in Exploration **2** with the one you created for the historic rainfall data in Activity **1**.

b. Consider a random number generator that generates integers from 0 to 12. Why shouldn't this simulation be used to model the probabilities associated with the number of heads when tossing 12 coins?

c. By examining your data from the exploration, what would you consider a "typical" interval of precipitation for May?

d. Do you think that the simulation from the exploration provides a reasonable model of Bernoulli's historic May precipitation? Explain your response.

Assignment

2.1 **a.** In Exploration **2**, what percentage of the coin-tossing data fell

1. within 1 standard deviation of the mean?
2. within 2 standard deviations of the mean?

b. In Activity **1,** Discussion Part **f**, you found that 67% of the May rainfall intervals were within 1 standard deviation of the mean and 96% were within 2 standard deviations of the mean. How do these percentages compare with your responses in Part **a**?

2.2 **a.** Use the information you recorded in Table **7** to answer the following questions.

1. What is the probability that the precipitation is between 10 cm and 12 cm?
2. What is the probability that the precipitation is greater than or equal to 16 cm?
3. What is the probability that the precipitation is less than 8 cm?

b. In Activity **1**, Problem **1.1**, you found that the historic probability of Bernoulli receiving between 10 cm and 12 cm of rainfall was 20%. The probability of receiving greater than 16 cm of rainfall was 27%, while the probability of receiving less than 8 cm of rainfall was 4%.

1. How do these values compare to your responses to Part **a**?
2. Discuss some possible explanations for any similarities or differences you observe.

2.3 An experiment that involved tossing 3 fair coins and counting the numbers of heads was repeated 50 times. The table below shows the relative frequencies that were observed.

No. of Heads	Frequency	Relative Frequency	Probability
0		0.12	
1		0.42	
2		0.28	
3		0.18	

a. What are the corresponding probabilities?

b. What are the corresponding frequencies?

c. What are the mean and the median of the data?

d. What values would you expect the probabilities to approach for a large number of simulations? Explain your response.

e. What is the expected value, in number of heads, for the experiment?

f. **1.** What is the mean of the binomial distribution for an experiment that consists of tossing a fair coin three times and counting the number of heads that appear?

2. How does this mean compare with the expected value calculated in Part **e**?

* * * * *

2.4 In major league baseball's World Series, the first team to win 4 games is declared the champion. Based on historic data, the probability that the World Series will last 4 games is 0.120, the probability that it will last 5 games is 0.253, the probability that it will last 6 games is 0.217, and the probability that it will last 7 games is 0.410.

a. Design a simulation to model the next 50 World Series.

b. Determine the mean and standard deviation for the data generated by the simulation.

c. Compare the mean and standard deviation of the data to the historic values.

* * * * * * * * * * *

Activity 3

Although your simulation appears to produce rainfall amounts similar to those in Bernoulli's historic rainfall data, some citizens are still concerned about the model. One critic argues that the simulation generates only 13 different outcomes, while the real amounts of Bernoulli's May precipitation can take on an infinite number of outcomes in the interval from 0 cm to 26 cm.

To address this criticism, you decide to modify the simulation by treating the probability distribution representing amounts of rainfall as a **continuous probability distribution**.

Discussion 1

a. What is the sum of all the relative frequencies in the coin-tossing simulation? Explain your response.

b. If all 13 intervals of precipitation were equally likely, how could you determine a relative frequency for each interval?

c. How many different amounts of precipitation can there be in each interval?

Mathematics Note

A **continuous probability distribution** results when the outcomes of an experiment can take on all possible real-number values within an interval.

In this situation, the probabilities of these outcomes can be represented graphically by the area enclosed by the x-axis, a specific real-number interval, and a distribution curve. The sum of the non-overlapping areas that cover the entire interval is 1.

Figure **3** shows one example of a continuous probability distribution. The probability that a value falls in the interval [0.5, 1.5] is equal to the area of region A, or 0.55. The probability that a value falls in the interval [1.5, 2.5] is equal to $1 - 0.55 = 0.45$, or the area of region B.

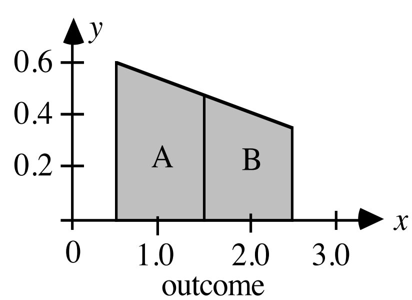


Figure 3: A continuous probability distribution

A **uniform probability distribution** is a continuous probability distribution in which all the probabilities over intervals of equal length are equal.

For example, consider the distribution of real numbers in the interval [0, 10], as shown in Figure **4**. The probability that a randomly selected real number will fall in the interval [4, 6] is $2 \bullet 0.1 = 0.2$ or 20%. Since this is a uniform probability distribution, the probability that a randomly selected real number will fall in any interval with a length of 2 is also 20%.

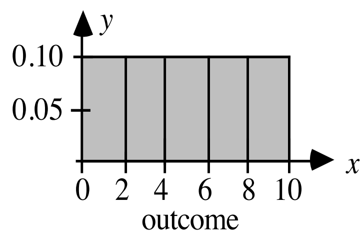


Figure 4: A uniform probability distribution

d. Why must the height of the uniform probability distribution in Figure **4** be 0.1?

e. Given an interval of real numbers, describe the height of its corresponding uniform probability distribution.

Exploration

a. The random number generator on many calculators produces numbers between 0 and 1. Assuming every number between 0 and 1 is equally likely to occur, draw a graph that you think represents the probability distribution of the possible outcomes.

b. Divide the numbers between 0 and 1 into intervals of equal lengths. Mark these intervals on your graph from Part **a**.

c. Select one interval on your graph. Shade the area that represents the probability of generating a random number that falls in this interval. Determine the area of the shaded region.

d. Generate 200 random numbers between 0 and 1. Record how many of these numbers fall in the interval you selected in Part **c**.

e. Determine the percentage of numbers generated that fell in the interval you selected in Part **c**. Compare this to the area you calculated in Part **c**.

f. One of the intervals in the rainfall simulation in Activity **2** represents rainfall values greater than or equal to 10 cm but less than 12 cm. Draw the probability distribution for the outcomes of a random number generator that produces numbers in the interval [10,12). Assume that all numbers in this interval are equally likely to occur.

Discussion 2

a. Does the experiment you conducted in Part **d** of the exploration represent a binomial experiment? Explain your response.

b. A uniform probability distribution, like all continuous probability distributions, shows the probability of all possible outcomes of an experiment. Given this fact, what is the total area in a uniform probability distribution?

c. How could you determine the probability that a randomly selected number between 0 and 1 will fall in the interval $[0.4, 0.5)$?

d. What is the probability that a randomly selected number between 0 and 1 will fall in each of the following intervals?

1. $[0.4, 0.45)$
2. $[0.4, 0.41)$
3. $[0.4, 0.401)$
4. $[0.4, 0.4001)$

e.

1. If the pattern in your responses to Part **d** above continues, what would be the probability of obtaining exactly 0.4?
2. What do you think is the probability of obtaining a specific number in any continuous probability distribution? Explain your response.

f.

1. Does the probability distribution shown in Figure **5** below appear to represent a continuous distribution? Explain your response.
2. Is this distribution uniform? Explain your response.

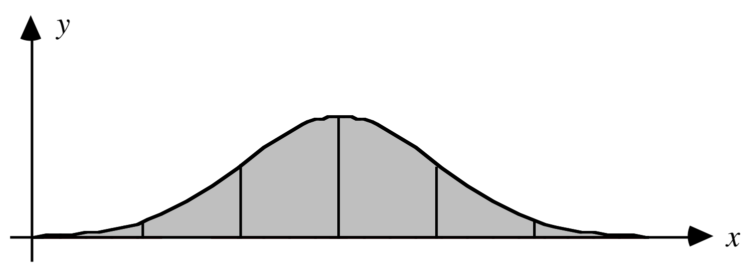


Figure 5: A probability distribution

Assignment

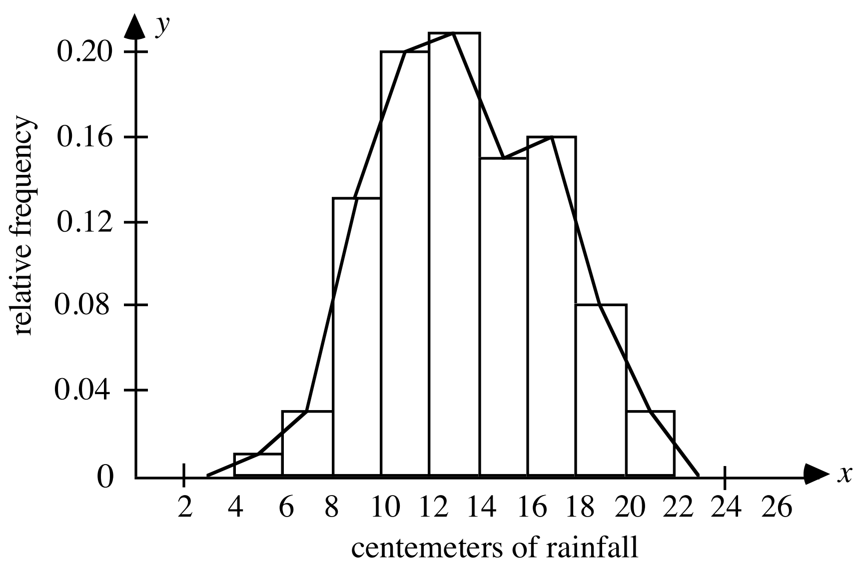
3.1 Given a continuous probability distribution, explain why it makes no difference whether or not an interval's endpoints are included when calculating the probability that a randomly selected value will fall in the interval.

3.2
a. Assume that a random number generator produces numbers in the open interval (–3, 3) so that every number in the interval is equally likely to occur. Draw a graph showing the probability distribution of the possible outcomes.

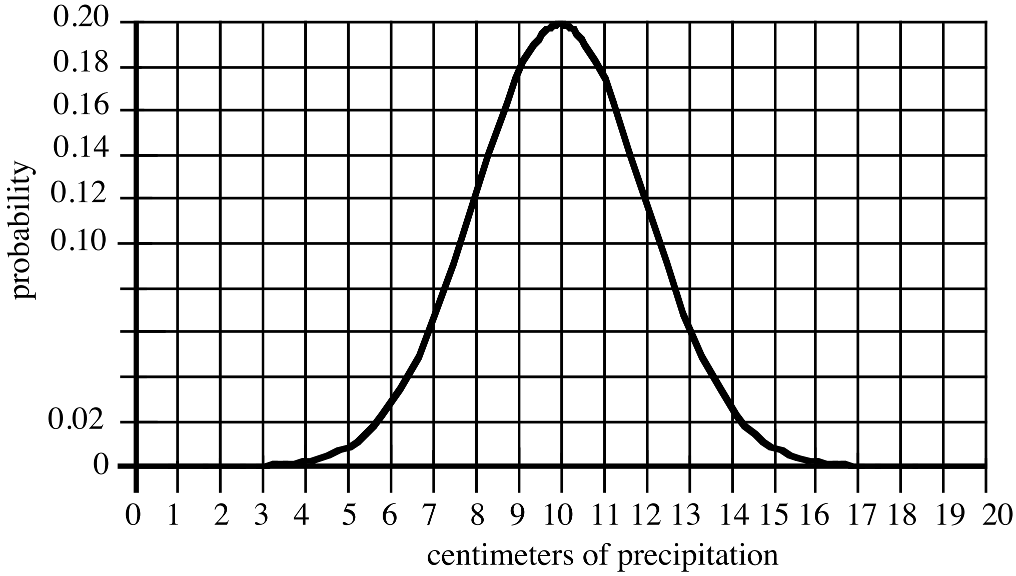
b. Using the random number generator described in Part **a**, what is the probability of obtaining a number in the interval (–1, 1)? Illustrate this probability by shading the appropriate portion of your graph.

c.
1. Create the random number generator described in Part **a**.
2. Generate 200 numbers in the interval (–3, 3).
3. Determine the percentage of those numbers that fall in the interval (–1, 1).
4. Compare this percentage to your response to Part **b**.

3.3 The relative frequency polygon for the precipitation data in Table **1** is shown below. Does this polygon represent a uniform probability distribution? Explain your response.



3.4 In the year 2093, the city of Hypatia has become the largest domed city in the western hemisphere. Hypatia's climate-control officer uses the following continuous probability distribution to determine the May precipitation totals for the city.



a. Determine the area under the curve in this probability distribution. Defend your response.

b. What is the mean amount of precipitation in May for Hypatia? Justify your response.

c. Use the graph to estimate the probability that the total precipitation in a given May will be at least 6 cm but less than 14 cm.

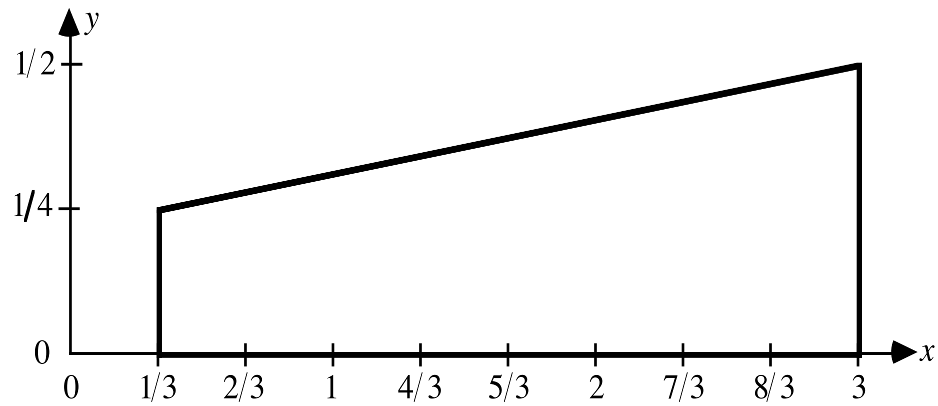
d. Hypatia's climate-control officer reports that the standard deviation of May rainfall is 2 cm. Use the graph to estimate each of the following:

1. the probability that the precipitation in a given May will be within 1 standard deviation of the mean

2. the probability that the precipitation in a given May will be within 2 standard deviations of the mean.

* * * * *

3.5 The polygon in the graph below is determined by the intersection of the lines $-3x + 32y = 7$, $x = 1/3$, $x = 3$, and the x-axis.



a. Explain why this polygon and its interior could represent a continuous probability distribution.

b. Could this polygon and its interior represent a uniform probability distribution? Explain your response.

c. Using this distribution, determine the probability that a randomly selected x-value in the interval $[1/3, 3]$ will be:

1. exactly 2
2. less than 2
3. greater than 2.

3.6 **a.** Assuming that a random number generator has a uniform probability distribution for numbers between 0 and 1, what is the probability that a randomly generated number will fall between:

1. 0.6 and 0.8?
2. 0.69 and 0.71?
3. 0.69999 and 0.70001?

b. What do your responses to Part **a** suggest about the probability of obtaining the number 0.7 from the random number generator?

* * * * * * * * * *

Activity 4

The rainfall model you created using a coin-tossing simulation is a discrete probability distribution. However, there is also a continuous probability distribution that can allow you to model Bernoulli's historic May rainfall: the **normal distribution**.

Mathematics Note

A **normal distribution** is a continuous probability distribution. As shown in Figure **6**, the graph of a normal distribution is symmetric about the mean and tapers to the left and right like a bell. The curve that describes the shape of the graph is the **normal curve**. The equation of the normal curve that models a particular set of data depends on the mean and standard deviation of the data.

As in all continuous probability distributions, the total area between the x-axis and a normal curve is 1. Approximately 68% of this area falls within 1 standard deviation of the mean, 95% within 2 standard deviations of the mean, and 99.7% within 3 standard deviations of the mean. This is the **68–95–99.7 rule**.

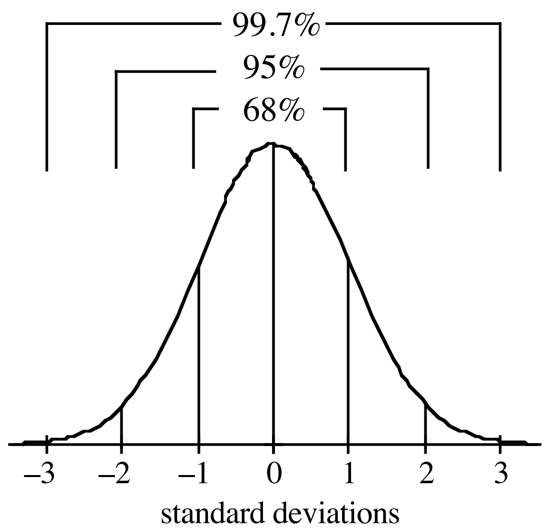


Figure 6: A normal curve and the 68–95–99.7 rule

Normal distributions can be used to model a wide variety of data sets. When this distribution provides a reasonable model, the 68–95–99.7 rule can help you characterize a population. For example, if a population appears to be normally distributed with a mean of 100 and a standard deviation of 10, then you would expect about 68% of the observations to fall between 90 and 110, 95% of the observations to fall between 80 and 120, and 99.7% of the observations to fall between 70 and 130.

Exploration

In Activity **2**, you found that you could approximate the shape of the probability distribution of Bernoulli's historic rainfall data by tossing 12 coins. In this exploration, you investigate the use of a normal curve to model the same data.

a. Obtain a copy of a program that calculates the theoretical probabilities of the outcomes of tossing coins from your teacher. This program plots the midpoints of the tops of the bars in a relative frequency histogram as a scatterplot. Each point in the scatterplot represents the theoretical probability of obtaining a specific number of heads. Based on the mean and standard deviation of the distribution, the program then draws a normal curve to model the data.

1. Use the program to draw the scatterplot and the normal curve for tossing 12 fair coins.
2. Record the mean and standard deviation of the distribution and describe the height and width of the corresponding normal curve.
3. Compare the shape of the curve to the scatterplot.

b. Repeat Part **a** for tossing 20 coins. Compare the mean, standard deviation, and maximum height of the normal curve to the values you obtained in Part **a**.

c. Repeat Part **a** for another value between 20 and 50 coins. Compare the mean, standard deviation, and maximum height of the normal curve to the values you obtained in Parts **a** and **b**.

Discussion

a. Why does a normal curve based on tossing 12 coins appear to be a good model for precipitation amounts in New Bernoulli?

b. How does increasing the number of coins tossed affect the mean and standard deviation of the distribution?

c. How does increasing the number of coins tossed affect the shape of the corresponding normal curve?

d. Using your knowledge of normal distributions, defend the following statement: "As the number of trials in a binomial experiment increases, the resulting probability distribution becomes approximately a normal distribution."

e. The normal curve shown in Figure **7** represents a probability distribution with a mean of 6 and a standard deviation of 1.2.

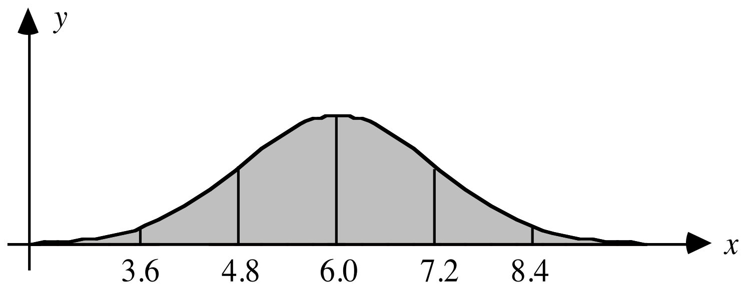


Figure 7: A normal distribution

Using this graph, how could you determine the probability of obtaining each of the following:

1. an x-value between 3.6 and 8.4?

2. an x-value greater than 8.4?

3. an x-value less than 7.2?

Assignment

4.1 In New Bernoulli, less than 3 cm of rain in a month can cause drought, while more than 20.5 cm of rain in a month can trigger floods. To reduce the possibility that the city will experience either drought or flood, the city council has asked you to use a normal distribution with a mean of 13.5 cm and a standard deviation of 3.5 cm to determine precipitation levels for the month of May.

a. Sketch a graph of a normal curve that represents this request. Indicate the mean and standard deviations on the horizontal axis.

b. If you follow the council's recommendations, what is the approximate probability that the May rainfall will be:

1. less than 3 cm?

2. more than 20.5 cm?

3. between 10 and 17 cm?

c. Write a short letter to the city council describing the chances of drought or flood during the month of May. Use probabilities and a discussion of the normal distribution to justify your report.

4.2 The average annual rainfall in a tropical forest is 120 cm, with a standard deviation of 10 cm. If amounts of annual rainfall are normally distributed, what is the probability that the forest will have less than 100 cm of rain this year?

4.3 The city of Pascal has an average annual rainfall of 40 cm, with a standard deviation of 5 cm. The city of Poisson has the same average annual rainfall, with a standard deviation of 2 cm. Levels of annual rainfall for both cities are normally distributed.

a. On the same set of axes, make a sketch of the distribution of rainfall amounts for each town.

b. The mayor of Poisson claims that the annual rainfall in her town is more often between 38 cm and 42 cm than the annual rainfall in Pascal. Is she right? Explain your response.

c. Describe the differences between the two graphs you sketched in Part **a**. Explain why these differences exist.

d. Describe the annual rainfall a resident of each town might expect.

* * * * *

4.4 Imagine that you are conducting an experiment using coins that are not fair. The probability of obtaining a head on any toss of one of these coins is 0.4. Use the program from the exploration in this activity to help you complete this problem.

a. What are the mean and standard deviation of the probability distribution of the number of heads when tossing:

1. 20 coins?
2. 40 coins?
3. 60 coins?

b. Are the formulas for the mean and standard deviation of a binomial distribution appropriate in this situation? Explain your response.

c. Repeat Part **a** when the probability of obtaining a head on any toss is 0.2.

d. Write a summary of your results to Parts **a–c**.

4.5 Kyla and Jon are playing a game in which players take turns rolling two six-sided dice, then moving tokens on a board. The number of squares each player moves is determined by the sum of the faces showing on the two dice.

a. Create a frequency distribution for the sum of the two dice.

b. Determine the mean, median, and standard deviation of the data in Part **a**.

c. Create a theoretical probability distribution for the sum of two dice.

d. Create a relative frequency histogram for the data in Part **c**.

e. Determine the probability that the sum of the faces showing on any given roll are within:

1. 1 standard deviation of the mean

2. 2 standard deviations of the mean

3. 3 standard deviations of the mean.

f. Do you think that the data could be modeled by a normal distribution? Explain your response.

g. On three consecutive turns, Kyla rolls 10 for the sum of the two dice. What is the probability of this occurrence?

* * * * * * * * * *

Research Project

A binomial distribution is also referred to as a **Bernoulli distribution**. During a period of more than two centuries, the Bernoulli family produced many prominent scientists and mathematicians. Research the Bernoullis and their accomplishments. Write a report on their contributions to the study of the normal curve and binomial distributions.

Summary Assessment

One of the radio stations in New Bernoulli offers 45 min of uninterrupted music several times each day. While planning the schedule for next week, the station manager asks you to collect some information on the lengths of popular songs.

a. Collect data on the lengths of at least 50 songs.

b. Assign interval lengths, find the relative frequencies, and construct a histogram of the data.

c. Based on your data, do the lengths of popular songs appear to be normally distributed? Explain your response.

d. Calculate the standard deviation and use it to describe the spread in the data.

e. How many songs would you expect the radio station to be able to play in 45 min? Explain your reasoning.

Module Summary

- The **frequency** of an item in a data set is the number of times that item is observed.
- The **relative frequency** of an item is the ratio of its frequency to the total number of observations in the data set.
- A **relative frequency table** includes columns that describe a data item or interval, its frequency, and its relative frequency.
- A **frequency histogram** consists of bars of equal width whose heights indicate the frequencies of intervals.
- A **frequency polygon** is formed by the line graph that connects the set of points (x,y), where x is the midpoint of an interval and y is the frequency of the interval. The base of a frequency polygon is the x-axis. If the frequency of the interval containing the greatest or least x-value is not 0, a vertex must be added to complete the polygon.
- In a **relative frequency histogram**, the heights of the bars represent relative frequencies instead of frequencies.
- In a **relative frequency polygon**, the y-values represent relative frequencies instead of frequencies.
- A **probability distribution** is the assignment of probabilities to a specific characteristic that belongs to each possible outcome of an experiment. If the set of outcomes is either finite or can have a one-to-one correspondence with the natural numbers, the distribution is **discrete**.
- A **binomial experiment** has the following characteristics:
  - It consists of a fixed number of repetitions of the same action. Each repetition is a **trial**.
  - The trials are independent of each other. In other words, the result of one trial does not influence the result of any other trial in the experiment.
  - Each trial has only two possible outcomes: a success or a failure.
  - The probability of a success remains the same from trial to trial.
  - The total number of successes is observed.
- A **binomial distribution** is the probability distribution for a binomial experiment.

- The mean of a binomial distribution is the product of the number of trials and the probability of a success. In other words, the mean μ can be found as follows, where n is the number of trials and p is the probability of a success:

$$\mu = np$$

- The standard deviation σ of a binomial distribution is the square root of the product of the number of trials, the probability of a success, and the probability of a failure:

$$\sigma = \sqrt{np(1-p)}$$

- A **continuous probability distribution** results when the outcomes of an experiment can take on all possible real-number values within an interval.
- A **uniform probability distribution** is a continuous probability distribution in which all the probabilities over intervals of equal width are equal.
- A **normal distribution** is a continuous probability distribution that is symmetric about the mean and tapers to the left and right like a bell. The curve that describes the shape of the graph is the **normal curve**. The equation of the normal curve that models a particular set of data depends on the mean and standard deviation of the data.
- As in all continuous probability distributions, the total area between the x-axis and a normal curve is 1. Approximately 68% of this area falls within 1 standard deviation of the mean, 95% within 2 standard deviations of the mean, and 99.7% within 3 standard deviations of the mean. This is the **68–95–99.7 rule**.

Selected References

Freedman, D., R. Pisani, and R. Purves. *Statistics*. New York: W. W. Norton & Co., 1980.

Information Please Almanac: Atlas and Yearbook 1993. Boston: Houghton Mifflin, 1993.

Moore, D. S., and G. P. McCabe. *Introduction to the Practice of Statistics*. New York: W. H. Freeman and Co., 1989.

National Climatic Data Center. *Climatological Data Annual Summary Montana*. Vols. 43–94. Asheville, NC: National Oceanic and Atmospheric Administration, 1992.