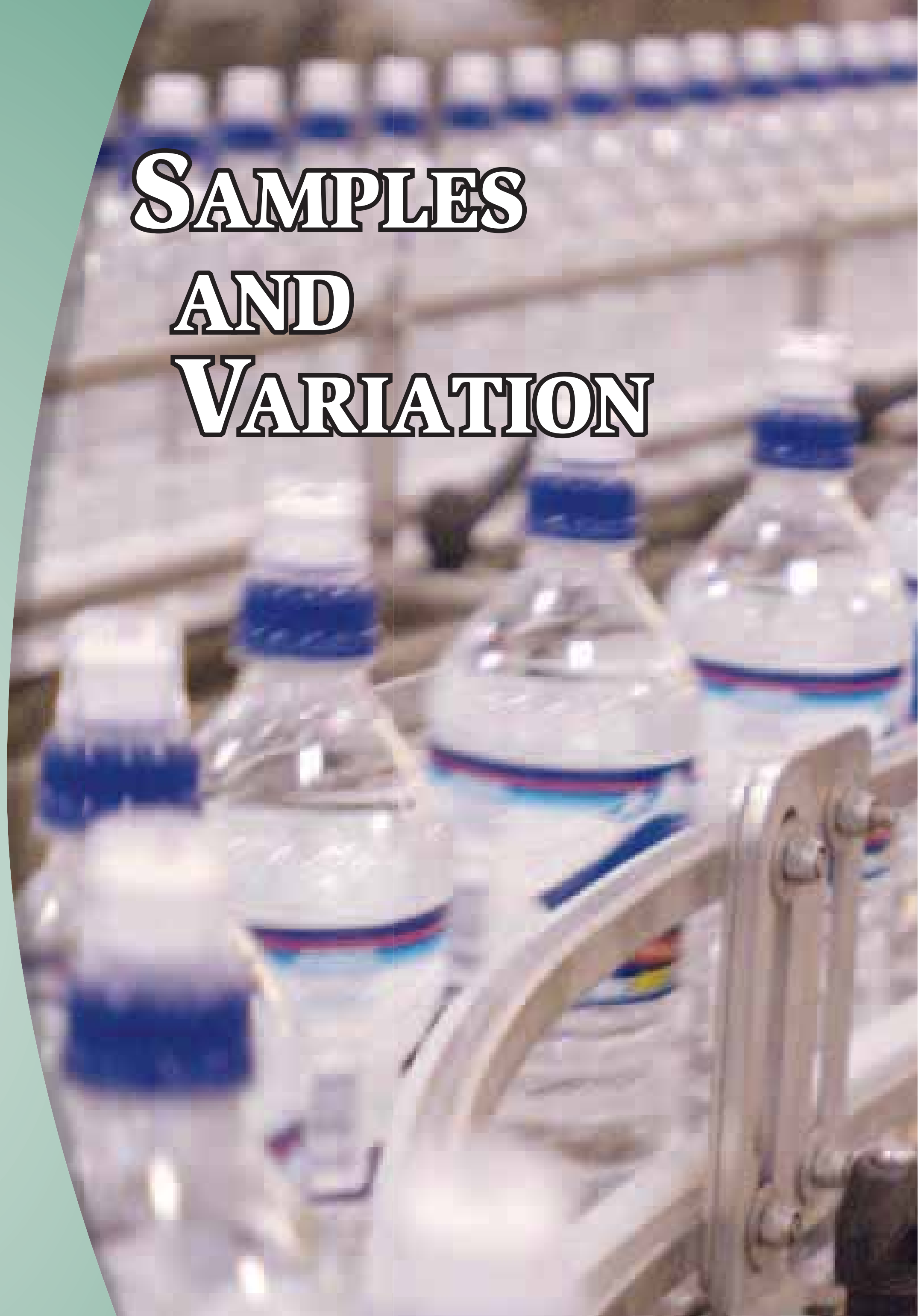
UNIT 4

Samples and Variation



Taking political polls and manufacturing car parts do not seem similar at first glance. However, both involve processes that have variation in the outcomes. When Gallup takes a poll, different samples of voters would give slightly different estimates of the president's popularity. When Ford Motor Company builds a car, body panels will vary slightly in their dimensions even if they are made by the same machine.

In this unit, you will investigate how understanding this variability helps both pollsters and manufacturers improve their products. The statistical methods they use can be effectively applied to any area in which there is variation in the process, essentially every area of human endeavor. The essential knowledge and skills required for this work are developed in three lessons of this unit.

Lessons

1 *Normal Distributions*

Describe characteristics of a normal distribution, compute and interpret a z-score, and estimate probabilities of events that have a normal distribution.

2 *Binomial Distributions*

Construct a binomial distribution and predict whether it will be approximately normal, compute the mean and standard deviation of a binomial distribution, and identify rare events.

3 *Statistical Process Control*

Recognize when the mean and standard deviation change in a plot over time, use control charts and the tests for out-of-control behavior, understand why it is best to watch a process for awhile before trying to adjust it, compute the probability of a false alarm, and use the Central Limit Theorem.

LESSON 1

Normal Distributions

Jet aircraft, like Boeing's latest 787 Dreamliner, are assembled using many different components. Parts for those components often come from other manufacturers from around the world. When different machines are used to manufacture the same part, the sizes of the produced parts will vary slightly. Even parts made by the same machine have slight variation in their dimensions.

Variation is inherent in the manufacturing of products whether they are made by computer-controlled machine or by hand. To better understand this phenomenon, suppose your class is working on a project making school award certificates. You have decided to outline the edge of a design on each certificate with one long piece of thin gold braid.

Think About This Situation

Consider the process of preparing the strip of gold braid for the certificate.

a What sources of variability would exist in the process?

b Individually, find the perimeter (in millimeters) of the edge of the design on a copy of the sample certificate.

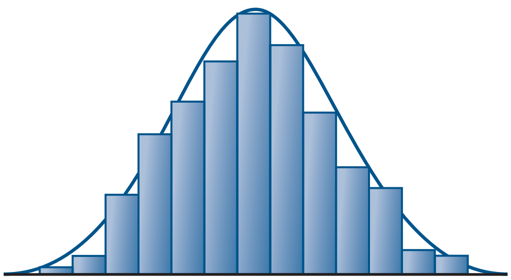
c Examine a histogram of the perimeters measured by the members of your class. Describe the shape, center, and spread of the distribution.

d What other summary statistics could you use to describe the center and spread?

In this lesson, you will explore connections between a normal distribution and its mean and standard deviation and how those ideas can be used in modeling the variability in common situations.

Investigation 1 Characteristics of a Normal Distribution

Many naturally occurring measurements, such as human heights or the lengths or weights of supposedly identical objects produced by machines, are **approximately normally distributed**. Their histograms are "bell-shaped," with the data clustered symmetrically about the mean and tapering off gradually on both ends, like the shape below.



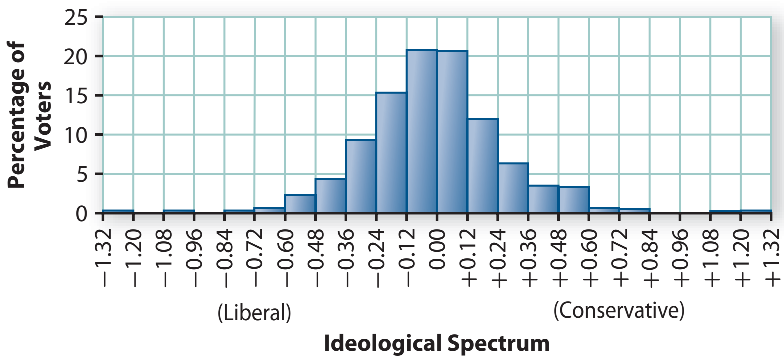
When measurements can be modeled by a distribution that is approximately normal in shape, the mean and standard deviation often are used to summarize the distribution's center and variability. As you work on the problems in this investigation, look for answers to the following question:

How can you use the mean and standard deviation to help you locate a measurement in a normal distribution?

1 In Course 1, *Patterns in Data*, you estimated the mean of a distribution by finding the balance point of a histogram. You estimated the standard deviation of a normal distribution by finding the distance to the right of the mean and to the left of the mean that encloses the middle 68% (about two-thirds) of the values. Use this knowledge to help analyze the following situations.

a. The histogram below shows the political points of view of a sample of 1,271 voters in the United States. The voters were asked a series of questions to determine their political philosophy and then were rated on a scale from liberal to conservative. Estimate the mean and standard deviation of this distribution.

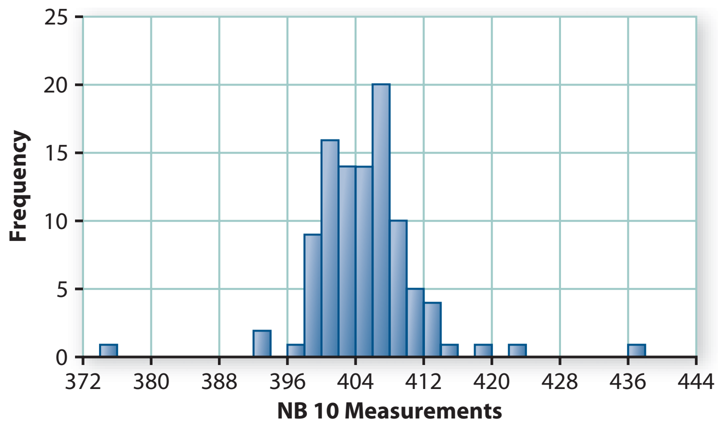
Political Philosophy



Source: Romer, Thomas, and Howard Rosenthal. 1984. Voting models and empirical evidence. *American Scientist*, 72: 465–473.

b. In a science class, you may have weighed something by balancing it on a scale against a standard weight. To be sure a standard weight is reasonably accurate, its manufacturer can have it weighed at the National Institute of Standards and Technology in Washington, D.C. The accuracy of the weighing procedure at the National Institute of Standards and Technology is itself checked about once a week by weighing a known 10-gram weight, NB 10. The histogram below is based on 100 consecutive measurements of the weight of NB 10 using the same apparatus and procedure. Shown is the distribution of weighings, in micrograms below 10 grams. (A microgram is a millionth of a gram.) Estimate the mean and standard deviation of this distribution.

NB 10 Weight

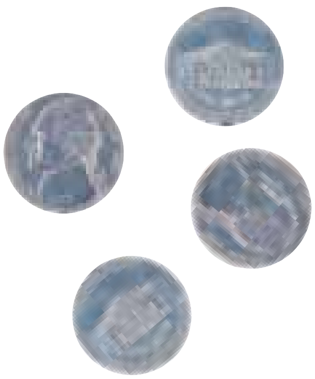


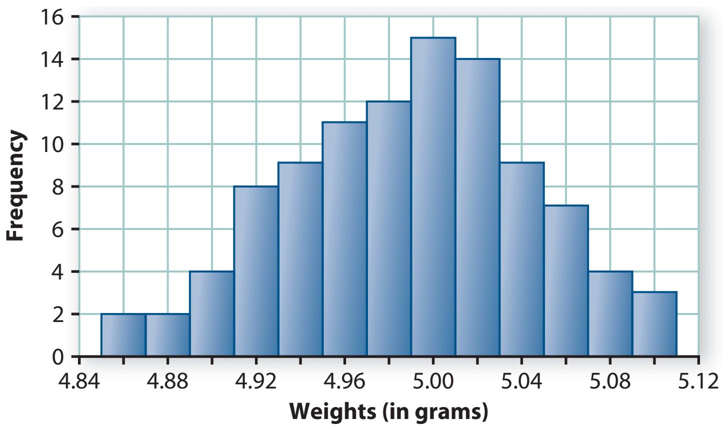
Source: Freedman, David, *et al. Statistics, 3rd edition*. New York: W. W. Norton & Co, 1998.

2 The data below and the accompanying histogram give the weights, to the nearest hundredth of a gram, of a sample of 100 new nickels.

Nickel Weights (in grams)

4.87	4.92	4.95	4.97	4.98	5.00	5.01	5.03	5.04	5.07
4.87	4.92	4.95	4.97	4.98	5.00	5.01	5.03	5.04	5.07
4.88	4.93	4.95	4.97	4.99	5.00	5.01	5.03	5.04	5.07
4.89	4.93	4.95	4.97	4.99	5.00	5.02	5.03	5.05	5.08
4.90	4.93	4.95	4.97	4.99	5.00	5.02	5.03	5.05	5.08
4.90	4.93	4.96	4.97	4.99	5.01	5.02	5.03	5.05	5.09
4.91	4.94	4.96	4.98	4.99	5.01	5.02	5.03	5.06	5.09
4.91	4.94	4.96	4.98	4.99	5.01	5.02	5.04	5.06	5.10
4.92	4.94	4.96	4.98	5.00	5.01	5.02	5.04	5.06	5.11
4.92	4.94	4.96	4.98	5.00	5.01	5.02	5.04	5.06	5.11





a. The mean weight of this sample is 4.9941 grams. Find the median weight from the table above. How does it compare to the mean weight?

b. Which of the following is the standard deviation?

0.0253 grams   0.0551 grams   0.253 grams   1 gram

c. On a copy of the histogram, mark points along the horizontal axis that correspond to the mean, one standard deviation above the mean, one standard deviation below the mean, two standard deviations above the mean, two standard deviations below the mean, three standard deviations above the mean, and three standard deviations below the mean.

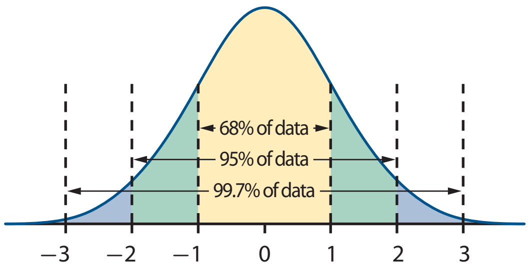
d. What percentage of the weights in the table above are within one standard deviation of the mean? Within two standard deviations? Within three standard deviations?

e. Suppose you weigh a randomly chosen nickel from this collection. Find the probability that its weight would be within two standard deviations of the mean.

In Problems 1 and 2, you looked at distributions of real data that had an *approximately* normal shape with mean $\overline{x}$ and standard deviation s. Each distribution was a sample taken from a larger population that is more nearly normal. In the rest of this investigation, you will think about theoretical populations that have a perfectly normal distribution.

The symbol for the *mean of a population* is μ, the lower-case Greek letter "mu." The symbol for the *standard deviation of a population* is σ, the lower case Greek letter "sigma." That is, use the symbol μ for the mean when you have an entire population or a theoretical distribution. Use the symbol $\overline{x}$ when you have a sample from a population. Similarly, use the symbol σ for the standard deviation of a population or of a theoretical distribution. Use the symbol s for the standard deviation of a sample.

All normal distributions have the same overall shape, differing only in their mean and standard deviation. Some look tall and skinny. Others look more spread out. All normal distributions, however, have certain characteristics in common. They are symmetric about the mean, 68% of the values lie within one standard deviation of the mean, 95% of the values lie within two standard deviations of the mean, and 99.7% of the values lie within three standard deviations of the mean.



3 Suppose that the distribution of the weights of newly minted coins is a normal distribution with mean μ of 5 grams and standard deviation σ of 0.10 grams.

a. Draw a sketch of this distribution. Then label the points on the horizontal axis that correspond to the mean, one standard deviation above and below the mean, two standard deviations above and below the mean, and three standard deviations above and below the mean.

b. Between what two values do the middle 68% of the weights of coins lie? The middle 95% of the weights? The middle 99.7% of the weights?

c. Illustrate your answers in Part b by shading appropriate regions in copies of your sketch.

4 Answer the following questions about normal distributions. Draw sketches illustrating your answers.

a. What percentage of the values in a normal distribution lie above the mean?

b. What percentage of the values in a normal distribution lie more than two standard deviations from the mean?

c. What percentage of the values in a normal distribution lie more than two standard deviations above the mean?

d. What percentage of the values in a normal distribution lie more than one standard deviation from the mean?

5 The weights of babies of a given age and gender are approximately normally distributed. This fact allows a doctor or nurse to use a baby's weight to find the weight percentile to which the child belongs. The table below gives information about the weights of six-month-old and twelve-month-old baby boys.

Weights of Baby Boys

	Weight at Six Months (in pounds)	Weight at Twelve Months (in pounds)
Mean μ	17.25	22.50
Standard Deviation σ	2.0	2.2

Source: Tannenbaum, Peter, and Robert Arnold. *Excursions in Modern Mathematics.* Englewood Cliffs, New Jersey: Prentice Hall. 1992.

a. On separate axes with the same scales, draw sketches that represent the distribution of weights for six-month-old boys and the distribution of weights for twelve-month-old boys. How do the distributions differ?

b. About what percentage of six-month-old boys weigh between 15.25 pounds and 19.25 pounds?

c. About what percentage of twelve-month-old boys weigh more than 26.9 pounds?

d. A twelve-month-old boy who weighs 24.7 pounds is at what percentile for weight? (Recall that a value x in a distribution lies at, say, the 27th **percentile** if 27% of the values in the distribution are less than or equal to x.)

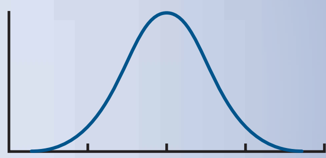
e. A six-month-old boy who weighs 21.25 pounds is at what percentile?

Summarize the Mathematics

In this investigation, you examined connections between a normal distibution and its mean and standard deviation.

a Describe and illustrate with sketches important characteristics of a normal distribution.

b On this graph, the distance between two adjacent tick marks on the horizontal axis is 5. Estimate the mean and the standard deviation of this normal distribution. Explain how you found your estimate.



Be prepared to share your ideas and reasoning with the class.

Check Your Understanding

Scores on the Critical Reading section of the SAT Reasoning Test are approximately normally distributed with mean μ of 502 and standard deviation σ of 113.

a. Sketch this distribution with a scale on the horizontal axis.

b. What percentage of students score between 389 and 615 on the verbal section of the SAT?

c. What percentage of students score over 615 on the verbal section of the SAT?

d. If you score 389 on the verbal section of the SAT, what is your percentile?

Investigation 2 Standardized Values

Often, you are interested in comparing values from two different distributions. For example: Is Sophie or her brother Pablo taller compared to others of their gender? Was your SAT score or your ACT score better? Answers to questions like these rely on your answer to the following question, which is the focus of this investigation:

How can you use standardized values to compare values from two different normal distributions?

1 Examine the table below, which gives information about the heights of young Americans aged 18 to 24. Each distribution is approximately normal.

Heights of American Young Adults (in inches)

	Men	Women
Mean μ	68.5	65.5
Standard Deviation σ	2.7	2.5

a. Sketch the two distributions. Include a scale on the horizontal axis.

b. Alexis is 3 standard deviations above average in height. How tall is she?

c. Marvin is 2.1 standard deviations below average in height. How tall is he?

d. Miguel is 74" tall. How many standard deviations above average height is he?

e. Jackie is 62" tall. How many standard deviations below average height is she?

f. Marina is 68" tall. Steve is 71" tall. Who is relatively taller for her or his gender, Marina or Steve? Explain your reasoning.

The **standardized value** tells how many standard deviations a given value lies from the mean of a distribution. For example, in Problem 1 Part b, Alexis is 3 standard deviations above average in height, so her standardized height is 3. Similarly, in Problem 1 Part c, Marvin is 2.1 standard deviations below average in height, so his standardized height is −2.1.

2 Look more generally at how standardized values are computed.

a. Refer to Problem 1, Parts d and e. Compute the standardized values for Miguel's height and for Jackie's height.

b. Write a formula for computing the standardized value z of a value x if you know the mean of the population μ and the standard deviation of the population σ.

3 Now consider how standardizing values can help you make comparisons. Refer to the table in Problem 1.

a. Find the standardized value for the height of a young woman who is 5 feet tall.

b. Find the standardized value for the height of a young man who is 5 feet 2 inches tall.

c. Is the young woman in Part a or the young man in Part b shorter, relative to his or her own gender? Explain your reasoning.

4 In an experiment about the effects of mental stress, subjects' systolic blood pressure and heart rate were measured before and after doing a stressful mental task. Their systolic blood pressure increased an average of 22.4 mm Hg (millimeters of Mercury) with a standard deviation of 2. Their heart rates increased an average of 7.6 beats per minute with a standard deviation of 0.7. Each distribution was approximately normal.

Suppose that after completing the task, Mario's blood pressure increased by 25 mm Hg and his heart rate increased by 9 beats per minute. On which measure did he increase the most, relative to the other participants? (Source: Mental Stress-Induced Increase in Blood Pressure Is Not Related to Baroreflex Sensitivity in Middle-Aged Healthy Men. *Hypertension*. 2000. vol. 35)

Summarize the Mathematics

In this investigation, you examined standardized values and their use.

a What information does a standardized value provide?

b What is the purpose of standardizing values?

c Kua earned a grade of 50 on a normally distributed test with mean 45 and standard deviation 10. On another normally distributed test with mean 70 and standard deviation 15, she earned a 78. On which of the two tests did she do better, relative to the others who took the tests? Explain your reasoning.

Be prepared to share your ideas and reasoning with the class.

Check Your Understanding

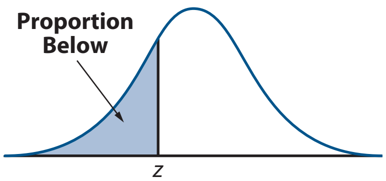
Refer to Problem 1 on page 243. Mischa Barton claims to be 5 feet 9 inches tall. Justin Timberlake claims to be 6 feet 1 inch tall. Who is taller compared to others of their gender, Mischa Barton or Justin Timberlake? Explain your reasoning.

Investigation 3 Using Standardized Values to Find Percentiles

In the first investigation, you were able to find the percentage of values within one, two, and three standard deviations of the mean. Your work in this investigation will extend to other numbers of standard deviations and help you answer the following question:

How can you use standardized values to find the location of a value in a distribution that is normal, or approximately so?

The following table gives the proportion of values in a normal distribution that are less than the given standardized value *z*. By standardizing values, you can use this table for any normal distribution. If the distribution is not normal, the percentages given in the table do not necessarily hold.



Proportion of Values in a Normal Distribution that Lie Below a Standardized Value *z*

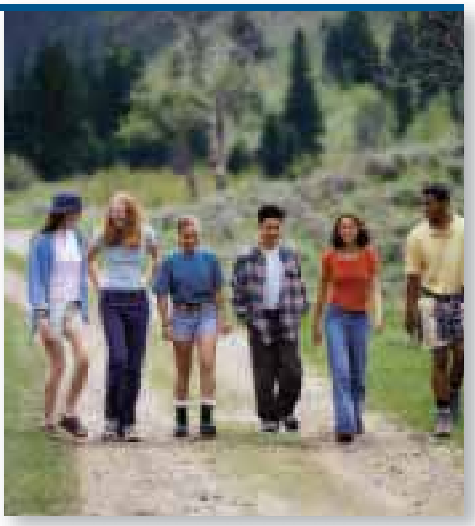
z	Proportion Below	*z*	Proportion Below	*z*	Proportion Below	*z*	Proportion Below
−3.5	0.0002	−1.7	0.0446	0.1	0.5398	1.9	0.9713
−3.4	0.0003	−1.6	0.0548	0.2	0.5793	2.0	0.9772
−3.3	0.0005	−1.5	0.0668	0.3	0.6179	2.1	0.9821
−3.2	0.0007	−1.4	0.0808	0.4	0.6554	2.2	0.9861
−3.1	0.0010	−1.3	0.0968	0.5	0.6915	2.3	0.9893
−3.0	0.0013	−1.2	0.1151	0.6	0.7257	2.4	0.9918
−2.9	0.0019	−1.1	0.1357	0.7	0.7580	2.5	0.9938
−2.8	0.0026	−1.0	0.1587	0.8	0.7881	2.6	0.9953
−2.7	0.0035	−0.9	0.1841	0.9	0.8159	2.7	0.9965
−2.6	0.0047	−0.8	0.2119	1.0	0.8413	2.8	0.9974
−2.5	0.0062	−0.7	0.2420	1.1	0.8643	2.9	0.9981
−2.4	0.0082	−0.6	0.2743	1.2	0.8849	3.0	0.9987
−2.3	0.0107	−0.5	0.3085	1.3	0.9032	3.1	0.9990
−2.2	0.0139	−0.4	0.3446	1.4	0.9192	3.2	0.9993
−2.1	0.0179	−0.3	0.3821	1.5	0.9332	3.3	0.9995
−2.0	0.0228	−0.2	0.4207	1.6	0.9452	3.4	0.9997
−1.9	0.0287	−0.1	0.4602	1.7	0.9554	3.5	0.9998
−1.8	0.0359	0.0	0.5000	1.8	0.9641		

1. Use the table to help answer the following questions. In each part, draw sketches that illustrate your answers.

   a. Suppose that a value from a normal distribution is two standard deviations below the mean. What proportion of the values are below it? What proportion are above it?

   b. If a value from a normal distribution is 1.3 standard deviations above the mean, what proportion of the values are below it? Above it?

   c. Based on the table, what proportion of values are within one standard deviation of the mean? Within two standard deviations of the mean? Within three standard deviations of the mean?



2. Reproduced below is the table of heights of Americans aged 18 to 24.

Heights of American Young Adults (in inches)

	Men	Women
Mean μ	68.5	65.5
Standard Deviation σ	2.7	2.5

   a. Miguel is 74 inches tall. What is his percentile for height? That is, what percentage of young men are the same height or shorter than Miguel?

   b. Jackie is 62 inches tall. What is her percentile for height?

   c. Abby is 5 feet 8 inches tall. What percentage of young women are between Jackie (Part b) and Abby in height?

   d. Gabriel is at the 90th percentile in height. What is his height?

   e. Yvette is at the 31st percentile in height. What is her height?

3. All 11th-grade students in Pennsylvania are tested in reading and math on the Pennsylvania System of School Assessment (PSSA). The mean score on the PSSA math test in 2006–2007 was 1,330 with standard deviation 253. You may assume the distribution of scores is approximately normal. (Source: www.pde.state.pa.us/a_and_t/cwp/view.asp?A=3&Q=129181)

   a. Draw a sketch of the distribution of these scores with a scale on the horizontal axis.

   b. What PSSA math score would be at the 50th percentile?

   c. What percentage of 11th graders scored above 1,500?

   d. Javier's PSSA score was at the 76th percentile. What was his score on this test?

Summarize the Mathematics

In this investigation, you explored the relation between standardized values and percentiles.

a How are standardized values used to find percentiles? When can you use this procedure?

b How can you find a person's score if you know his or her percentile and the mean and standard deviation of the normal distribution from which the score came?

Be prepared to share your ideas and reasoning with the class.

Check Your Understanding

Standardized values can often be used to make sense of scores on aptitude and intelligence tests.

a. Actress Brooke Shields reportedly scored 608 on the math section of the SAT. When she took the SAT, the scores were approximately normally distributed with an average on the math section of about 462 and a standard deviation of 100.

i. How many standard deviations above average was her score?

ii. What was Brooke Shields' percentile on the math section of the SAT?

b. The IQ scores on the Stanford-Binet intelligence test are approximately normal with mean 100 and standard deviation 15. Consider these lines from the movie *Forrest Gump*.

> **Mrs. Gump:** Remember what I told you, Forrest. You're no different than anybody else is. Did you hear what I said, Forrest? You're the same as everybody else. You are no different.
>
> **Principal:** Your boy's … different, Miz Gump. His IQ's 75.
>
> **Mrs. Gump:** Well, we're all different, Mr. Hancock.

i. How many standard deviations from the mean is Forrest's IQ score?

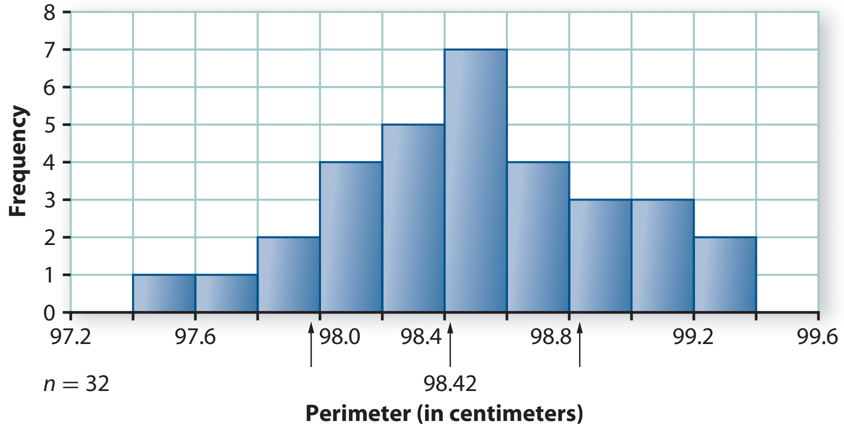
ii. What percentage of people have an IQ higher than Forrest's IQ?

On Your Own

Applications

1 Thirty-two students in a drafting class were asked to prepare a design with a perimeter of 98.4 cm. A histogram of the actual perimeters of their designs is displayed below. The mean perimeter was 98.42 cm.

Design Perimeters



a. What might explain the variation in perimeters of the designs?

b. The arrows mark the mean and the points one standard deviation above the mean and one standard deviation below the mean. Use the marked plot to estimate the standard deviation for the class' perimeters.

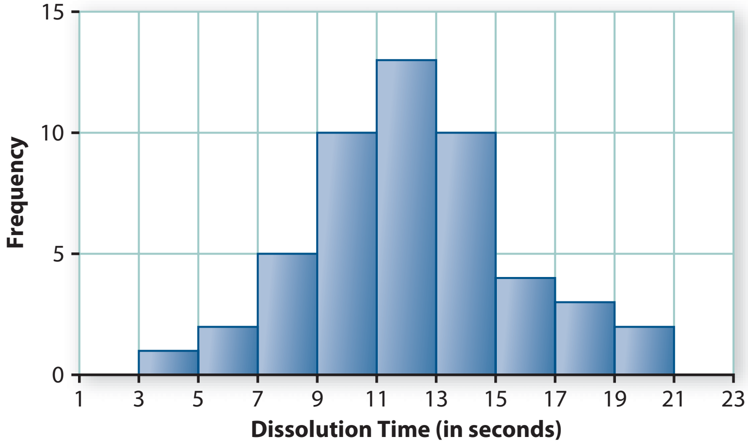
c. Estimate the percentage of the perimeters that are within one standard deviation of the mean. Within two standard deviations of the mean.

d. How do the percentages in Part c compare to the percentages you would expect from a normal distribution?

2 For a chemistry experiment, students measured the time for a solute to dissolve. The experiment was repeated 50 times. The results are shown in the chart and histogram at the top of the next page.

Dissolution Time (in seconds)

4	5	6	7	8	8	8	8	9	9
9	10	10	10	10	10	10	10	11	11
11	11	11	12	12	12	12	12	12	12
12	13	13	13	13	13	14	14	14	14
14	15	15	16	16	17	17	17	19	19



a. The mean time for the 50 experiments is 11.8 seconds. Find the median dissolution time from the table above. How does it compare to the mean dissolution time?

b. Which of the following is the best estimate of the standard deviation?

1.32 seconds   3.32 seconds   5.32 seconds

c. On a copy of the histogram, mark points along the horizontal axis that correspond to the mean, one standard deviation above the mean, one standard deviation below the mean, two standard deviations above the mean, two standard deviations below the mean, three standard deviations above the mean, and three standard deviations below the mean.

d. What percentage of the times are within one standard deviation of the mean? Within two standard deviations? Within three standard deviations?

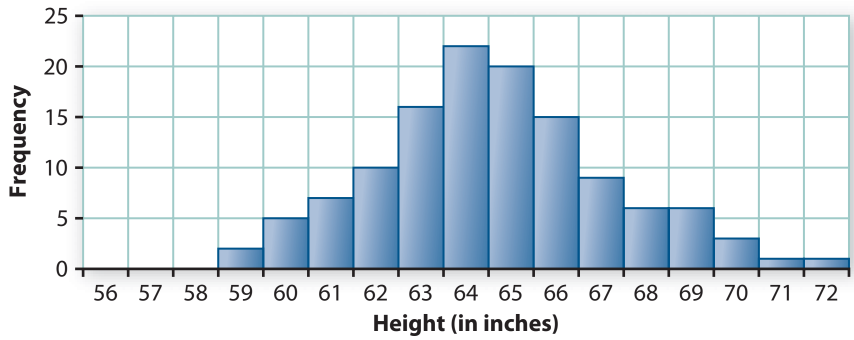
e. How do the percentages in Part d compare to the percentages you would expect from a normal distribution?

3 The table and histogram below give the heights of 123 women in a statistics class at Penn State University in the 1970s.

Female Students' Heights

Height (in inches)	Frequency	Height (in inches)	Frequency
59	2	66	15
60	5	67	9
61	7	68	6
62	10	69	6
63	16	70	3
64	22	71	1
65	20	72	1

Source: Joiner, Brian L. 1975. Living histograms. *International Statistical Review 3*: 339–340.



a. The mean height of the women in this sample is approximately 64.626 inches. Which of the following is the best estimate of the standard deviation?

0.2606 inches   0.5136 inches   2.606 inches   5.136 inches

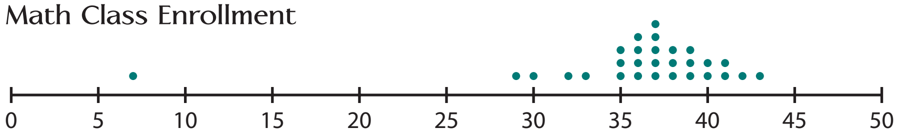
b. On a copy of the histogram, mark points along the horizontal axis that correspond to the mean, one standard deviation above the mean, one standard deviation below the mean, two standard deviations above the mean, two standard deviations below the mean, three standard deviations above the mean, and three standard deviations below the mean.

c. What percentage of the heights are within one standard deviation of the mean? Within two standard deviations? Within three standard deviations?

d. Suppose you pick a female student from the class at random. Find the probability that her height is within two standard deviations of the mean.

4 Suppose a large urban high school has 29 math classes. The number of students enrolled in each of the classes is displayed on the plot below. The outlier is a calculus class with only 7 students enrolled. One way to handle an outlier is to report a summary statistic computed both with and without it. If the two values are not very different, it is safe to report just the value that includes the outlier.

Math Class Enrollment



a. Justina computed the mean twice, with and without the outlier. The two means were 35.97 and 37.00. Which mean was computed with the outlier?

b. Justina also computed the standard deviation twice, with and without the outlier. The two standard deviations were 3.25 and 6.34. Which standard deviation was computed with the outlier?

c. How many standard deviations from the mean is the enrollment in the class with 43 students when the calculus class is included in computing the mean and standard deviation? When the calculus class is not included in the calculations?

d. Suppose five students drop out of each class. What will be the new mean and standard deviation if the calculus class is included in the computations?

5 Runners in the Boston Marathon compete in divisions determined by age and gender. In a recent marathon, the mean time for the 18- to 39-year-old women's division was 225.31 minutes with standard deviation 26.64 minutes. The mean time for the 50- to 59-year-old women's division was 242.58 minutes with standard deviation 21.78 minutes. In that marathon, a 34-year-old woman finished the race in 4:00:15, and a 57-year-old woman finished the race in 4:09:08 (hours:minutes:seconds).

a. How many standard deviations above the mean for her division was each runner?

b. Write a formula that gives the number of standard deviations from the mean for a time of x minutes by a woman in the 18- to 39-year-old division.

c. Write a formula that gives the number of standard deviations from the mean for a time of x minutes by a woman in the 50- to 59-year-old division.

6 The length of a human pregnancy is often said to be 9 months. Actually, the length of pregnancy from conception to natural birth varies according to a distribution that is approximately normal with mean 266 days and standard deviation 16 days.

a. Draw a sketch of the distribution of pregnancy lengths. Include a scale on the horizontal axis.

b. What percentage of pregnancies last less than 250 days?

c. What percentage of pregnancies are longer than 298 days?

d. To be in the shortest 2.5% of pregnancies, what is the longest that a pregnancy can last?

e. What is the median length of pregnancy?

7 Scores on the mathematics section of the SAT Reasoning Test are approximately normally distributed with mean 515 and standard deviation 114. Scores on the mathematics part of the ACT are approximately normally distributed with mean 21.0 and standard deviation 5.1.

a. Sketch graphs of the distribution of scores on each test. Include a scale on the horizontal axis.

b. What percentage of the SAT scores lie above 629? Above what ACT score would this same percentage of scores lie?

c. What ACT score is the equivalent of an SAT score of 450?

d. Find the percentile of a person who gets an SAT score of 450.

e. One of the colleges to which Eliza is applying accepts either SAT or ACT mathematics scores. Eliza scored 680 on the mathematics part of the SAT and 27 on the mathematics section of the ACT. Should she submit her SAT or ACT mathematics score to this college? Explain your reasoning.

8 Many body dimensions of adult males and females in the United States are approximately normally distributed. Approximate means and standard deviations for shoulder width are given in the table below.

U.S. Adult Shoulder Width (in inches)

	Men	Women
Mean μ	17.7	16.0
Standard Deviation σ	0.85	0.85

a. What percentage of women have a shoulder width of less than 15.5 inches? Of more than 15.5 inches?

b. What percentage of men have a shoulder width between 16 and 18 inches?

c. What percentage of American women have a shoulder width more than 17.7 inches, the average shoulder width for American men?

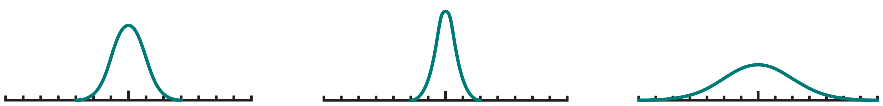
d. What percentage of men will be uncomfortable in an airplane seat designed for people with shoulder width less than 18.5 inches? What percentage of women will be uncomfortable?

e. If you sampled 100,000 men, approximately how many would you expect to be uncomfortable in an airline seat designed for people with shoulder width less than 18.5 inches? If you sampled 100,000 women, approximately how many would you expect to be uncomfortable in an airline seat designed for people with shoulder width less than 18.5 inches?

Connections

9 Three very large sets of data have approximately normal distributions, each with a mean of 10. Sketches of the overall shapes of the distributions are shown below. The scale on the horizontal axis is the same in each case. The standard deviation of the distribution in Figure A is 2. Estimate the standard deviations of the distributions in Figures B and C.

Figure A **Figure B** **Figure C**



10 How is the formula for the standard deviation $s = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n - 1}}$ like the distance formula?

11 If a set of data is the entire population you are interested in studying, you compute the population standard deviation using the formula:

$$\sigma = \sqrt{\frac{\Sigma(x - \mu)^2}{n}}$$

If you are looking at a set of data as a sample from a larger population of data, you compute the standard deviation using this formula:

$$s = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n - 1}}$$

a. For a given set of data, which is larger, σ or s? Explain your reasoning.

b. Does it make much difference whether you divide by n or by $n - 1$ if $n = 1{,}000$? If $n = 15$?

c. A sample tends to have less variability than the population from which it came. How does the formula for s, the standard deviation for a sample, account for this fact?

12 Suppose a normal distribution has mean 100 and standard deviation 15. Now suppose every value in the distribution is converted to a standardized value.

a. What is true about the mean of the standardized values?

b. What is true about the standard deviation of the standardized values?

13 In a normal distribution, about 5% of the values are more than two standard deviations from the mean. If a distribution is not approximately normal, is it possible to have more than 5% of the values two or more standard deviations from the mean? In this task, "standard deviation" refers to the population standard deviation σ, as defined in Task 11.

a. Make up two sets of numbers and compute the percentage that are two or more standard deviations from the mean. Your objective is to get the largest percentage that you can.

b. What is the largest percentage of numbers that you were able to find in Part a? Compare your results to those of other students completing this task.

14 The formula for Pearson's correlation that you studied in the Course 2 *Regression and Correlation* unit is

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right).$$

Here, n is the sample size, $\bar{x}$ is the mean of the values of x, $\bar{y}$ is the mean of the values of y, s_x is the standard deviation of the values of x, and s_y is the standard deviation of the values of y.

a. Use this formula to find the correlation between x and y for the following pairs of numbers.

x	*y*
1	1
2	2
3	6

b. Explain the meaning of the correlation in the context of standardized values.

Reflections

15 Why do we say that distributions of real data are "approximately" normally distributed rather than say they are normally distributed?

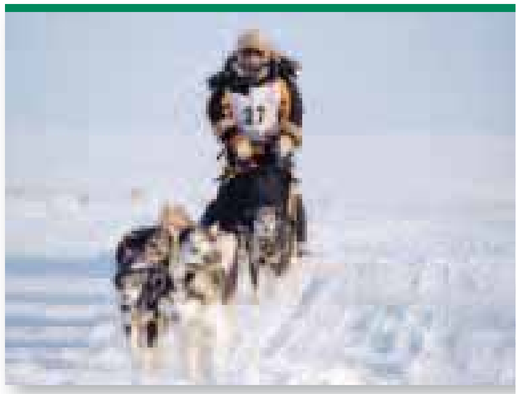
16 Under what conditions will the standard deviation of a data set be equal to 0? Explain your reasoning.

17 Consider the weights of the dogs in the following two groups.

- the dogs pulling a sled in a trans-Alaska dog sled race
- the dogs in a dog show, which includes various breeds of dogs

a. Which group would you expect to have the larger mean weight? Explain your reasoning.

b. Which group would you expect to have the larger standard deviation? Explain your reasoning.



18 Is it true that in all symmetric distributions, about 68% of the values are within one standard deviation of the mean? Give an example to illustrate your answer.

19 ACT and SAT scores have an approximately normal distribution. Scores on classroom tests are sometimes assumed to have an approximately normal distribution.

a. What do teachers mean when they say they "grade on a curve"?

b. Explain how a teacher might use a normal distribution to "grade on a curve."

c. Under what circumstances would you want to be "graded on a curve"?

Extensions

20 The producers of a movie did a survey of the ages of the people attending one screening of the movie. The data are shown in the table.

Saturday Night at the Movies

Age (in years)	Frequency
12	2
13	26
14	38
15	32
16	22
17	10
18	8
19	8
20	6
21	4
22	1
23	3
27	2
32	2
40	1

a. Compute the mean and standard deviation for this sample of ages. Do this without entering each of the individual ages into a calculator or computer software. (For example, do not enter the age "14" thirty-eight times).

b. In this distribution, what percentage of the values fall within one standard deviation of the mean? Within two standard deviations of the mean? Within three standard deviations of the mean?

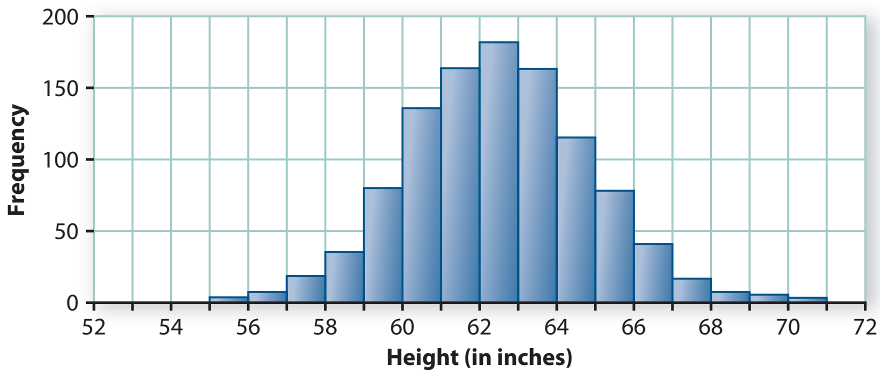
c. Compare the percentages from Part b to those from a normal distribution. Explain your findings in terms of the shapes of the two distributions.

21 In 1903, Karl Pearson and Alice Lee collected the heights of 1,052 mothers. Their data are summarized below. A mother who was exactly 53 inches tall was recorded in the 53–54 inches row.

Heights of Mothers (in inches)

Height	Number of Mothers	Height	Number of Mothers
52–53	1	62–63	183
53–54	1	63–64	163
54–55	1	64–65	115
55–56	2	65–66	78
56–57	7	66–67	41
57–58	18	67–68	16
58–59	34	68–69	7
59–60	80	69–70	5
60–61	135	70–71	2
61–62	163		

Source: Pearson, Karl and Alice Lee. 1903. On the laws of inheritance in man. *Biometrika*: 364.



a. Is this distribution of heights approximately normal? Why or why not?

b. Collect the heights of 30 mothers. How does the distribution of your sample compare to the distribution of heights from mothers in 1903?

c. What hypothesis might you make about heights of mothers today? Design a plan that you could use to test your hypothesis.

22 The equation of the curve that has the shape of a normal distribution is:

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

In this formula, μ is the mean of the normal distribution, σ is the standard deviation, and the number e is approximately equal to 2.71828.

a. Use your calculator or computer software to graph the normal curve that has a mean of 0 and a standard deviation of 1.

b. Describe what happens to the curve if you increase the mean. Describe what happens if you increase the standard deviation.

c. Normal curves have two "bends," or **points of inflection**. One is on the left side of the curve where the graph changes from curved up to curved down. The other is on the right side of the curve where the graph changes from curved down to curved up. Estimate the point where the "bend" seems to occur in the curve in Part a. What relation does this point have to the mean and standard deviation?

23 Discuss whether the situations below are consistent with what you know about normal distributions and IQ tests. Recall that the mean and standard deviation for IQ tests are $\mu = 100$ and $\sigma = 15$. What could account for any inconsistencies that you see?

a. One of the largest K–12 districts in the country, educates between 700,000 and 800,000 children. In this district, there are special magnet schools for "highly gifted" children. The only way for a child in this district to be classified as highly gifted is to score 145 or above on an IQ test given by a school psychologist. Recently, at one gifted magnet school, there were 61 students in the sixth-grade class.

b. One way for a child to be identified as gifted in California is to have an IQ of 130 or above. A few years ago, a Los Angeles high school had a total enrollment of 2,830 students, of whom 410 were identified as gifted.

Review

24 Suppose that the cost of filling up a car with gasoline is a function of the number of gallons and can be found using the rule $C(g) = 2.89g$.

a. What does the 2.89 tell you about this situation?

b. What is the cost of 4.2 gallons of gas?

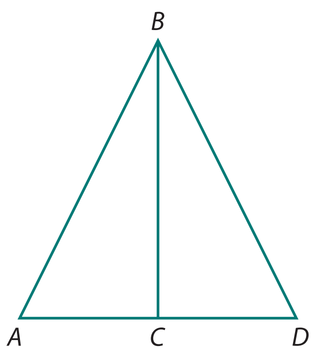
c. Find the value of g so that $C(g) = 19.65$. Then explain what your solution means in this context.

d. Is this relationship a direct variation, inverse variation, or neither? Explain.

25 In the diagram at the right, $\overline{BC}$ is the perpendicular bisector of $\overline{AD}$, and $AD = 20$ cm. If possible, complete each of the following tasks. If not possible, explain why not.

a. Prove that $\triangle ACB \cong \triangle DCB$.

b. Find the area of $\triangle ABD$.



26 Solve each inequality and graph the solution on a number line. Then describe the solution using interval notation.

a. $3(x + 7) \geq 5$

b. $10 - 8x < 4(8 - 4x)$

c. $2x \geq 5x - 3$

d. $x^2 > 4$

e. $x^2 - 5x < 0$

27 Rewrite each expression in factored form.

a. $3x^2 + 9x$

b. $2x^2 + 9x + 4$

c. $x^2 - 64$

d. $x^2 - 12x - 45$

28 Try to answer each of the following using mental computation only.

a. What is 25% of 400?

b. What is 60% of 80?

c. 20 is what percent of 50?

d. 180 is what percent of 270?

e. 6 is 20% of what number?

f. 14 is 2% of what number?

29 Given $\triangle XYZ$ with $m\angle X = 90°$ and $\sin Y = \frac{1}{2}$:

a. Find a set of possible exact side lengths for the triangle. Are your side lengths the only ones possible? If so, explain why. If not, explain why not.

b. Which of the following is the value for cos Z?

$$\frac{1}{2} \qquad \frac{2}{\sqrt{3}} \qquad \sqrt{3} \qquad \frac{\sqrt{3}}{2}$$

c. Find $m\angle Y$.

30 Rewrite each expression in simplest equivalent form.

a. $10 - 4(5x + 1)$

b. $(2x + 1)(3 - x) + 5(x + 3)$

c. $\frac{6x + 9}{3}$

d. $(4x + 5)^2$

31 Recall that two events A and B are independent if the occurrence of one of the events does not change the probability that the other occurs. For each situation, decide if it is reasonable to assume that the events are independent. Then find the probability of both events occurring.

a. Rolling a pair of dice one time

Event A: Getting doubles

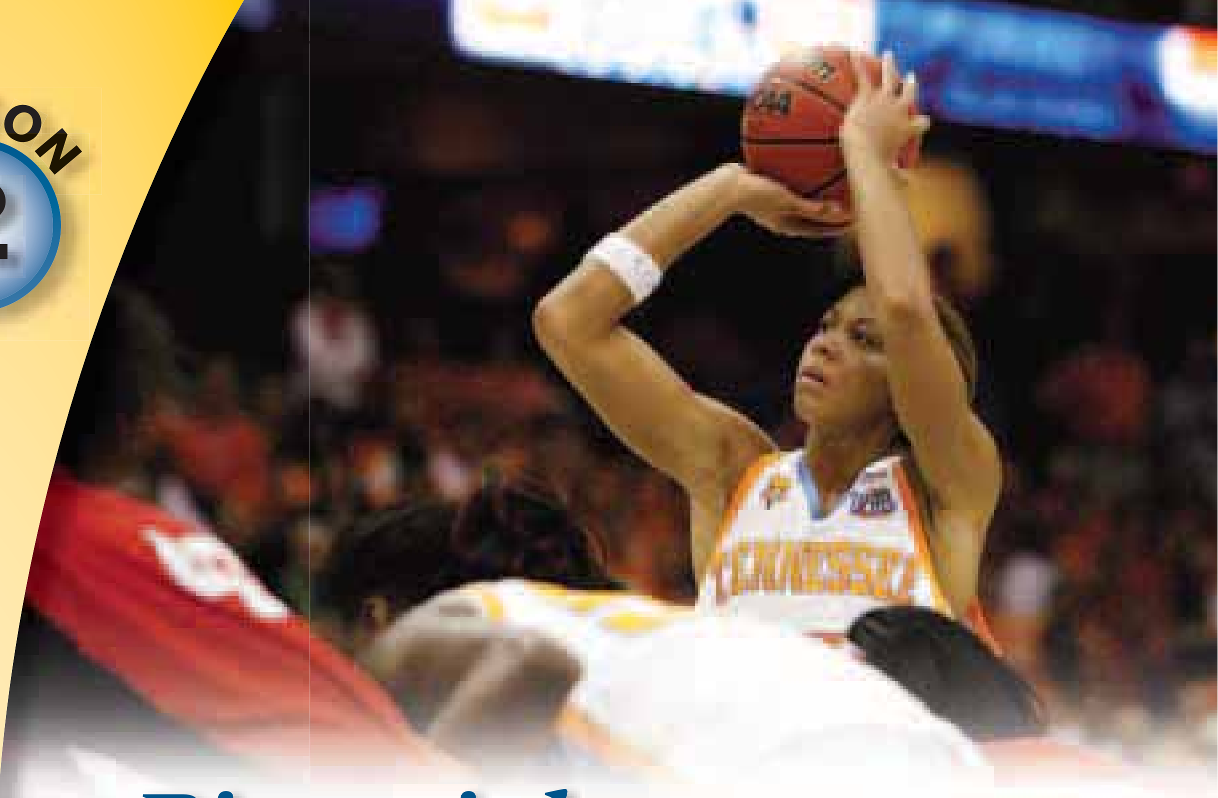
Event B: Getting a sum that is even

b. Flipping a coin twice

Event A: Getting heads on the first flip

Event B: Getting heads on the second flip

LESSON 2



Binomial Distributions

In the first lesson, you learned about properties of the normal distribution. This lesson is about binomial distributions. For example, suppose you flip a coin 10 times and count the number of heads. The binomial distribution for this situation gives you the probability of getting 0 heads, of getting 1 head, of getting 2 heads, and so on, up to the probability of getting 10 heads. A similar situation occurs when counting the number of successful free throw attempts for a specific basketball player in a game or season.

Consider the case of Candace Parker who played basketball for the University of Tennessee from 2005 to 2008. She is a versatile player usually playing forward, but she was listed on Tennessee's roster as forward, center, and guard. She played in 110 games and made 526 of her 738 attempted free throws. (Source: chicagosports.sportsdirectinc.com)

Think About This Situation

Suppose you were watching Candace Parker play basketball in 2008.

a What is your best estimate of the probability that Candace Parker will make a single free throw?

b How many free throws do you expect her to make in 20 attempts?

c Suppose Candace Parker attempts 20 free throws in a series of games. Would you be surprised if she made all 20? If she made only 12?

d How would you simulate 10 runs of the situation of Candace Parker attempting 20 free throws? What assumptions are you making that may be different from the real-life situation?

In this lesson, you will learn how to answer questions like those above by describing the shape, center, and spread of the distribution of the possible numbers of successes.

Investigation 1 Shape, Center, and Spread

Repeating the same process for a fixed number of *trials* and counting the number of "successes" is a common situation in probability. For example, you might roll a pair of dice 50 times and count the number of doubles. Or, you might survey 500 randomly selected U.S. teens and count the number who play soccer. (Recall that getting a random sample of 500 U.S. teens is equivalent to putting the names of all U.S. teens in a hat and drawing out 500 at random.) These are called **binomial situations** if they have the four characteristics listed below.

- There are two possible outcomes on each trial called "success" and "failure."
- Each trial is independent of the others. That is, knowing what happened on previous trials does not change the probability of a success on the next trial. (If you take a relatively small random sample from a much larger population, you can consider the trials independent.)
- There is a fixed number of trials n.
- The probability p of a success is the same on each trial.

Your work in this investigation will help you answer the following question:

What are the shape, mean, and standard deviation of a distribution of the number of successes in a binomial situation?

1 Determine whether each of the following situations is a binomial situation by identifying:

- the two possible outcomes on each trial
- whether the trials are independent
- the number of trials
- the probability of a success on each trial

a. You flip a coin 20 times and count the number of heads.

b. You roll a six-sided die 60 times and count the number of times you get a 2.

c. About 51% of the residents of the U.S. are female. You take a randomly selected sample of 1,200 U.S. residents and count the number of females. (Source: *Gender: 2000, Census 2000 Brief*, September 2001, page 1.)

d. In 2005, 60% of children lived in areas that did not meet one or more of the Primary National Ambient Air Quality Standards. You select 75 children at random from the U.S. and count the number who live in areas that do not meet one or more of these standards. (Source: *America's Children: Key National Indicators of Well-Being 2007*, page 30.)

e. The player with the highest career free throw percentage in NBA history is Mark Price who made 2,135 free throws out of 2,362 attempts. Suppose Mark Price shoots 20 free throws and you count the number of times he makes his shot. (Source: www.nba.com/statistics/)

CPMP-Tools

Recall that in probability, **expected value** means the *long-run average* or *mean* value. For example, if you flip a coin 7 times, you might get 3 heads, you might get 4 heads, and you might get more or fewer. Over many sets of 7 flips, however, the average or **expected number** of heads will be 3.5. So, you can say that you *expect* to get 3.5 heads if you flip a coin 7 times.

2 In the Course 2 *Probability Distributions* unit, you learned that if the probability of a success on each trial of a binomial situation is p, then the expected number of successes in n trials is np.

a. What formula gives you the expected number of failures?

b. For the binomial situations in Problem 1, what are the expected number of successes and the expected number of failures?

You can use statistics software or a command on your calculator to simulate a binomial situation. For example, suppose you want to simulate flipping a coin 100 times and counting the number of heads. From the calculator Probability menu, select **randBin(**. Type in the number of trials and the probability of a success, **randBin(100,.5)** and press ENTER. The calculator returns the number of successes in 100 trials when the probability of a success is 0.5. The "Random Binomial" feature of simulation software like in *CPMP-Tools* operates in a similar manner.

3 Give the calculator or software command that you would use to simulate each of the binomial situations in Problem 1. Then do one run and record the number of successes.

4 Suppose you flip a coin 100 times and count the number of heads.

a. What is the expected number of heads?

b. If everyone in your class flipped a coin 100 times and counted the number of heads, how much variability do you think there would be in the results?

c. Use the **randBin** function of your calculator or the "Random Binomial" feature of simulation software to simulate flipping a coin 100 times. Record the number of heads. Compare results with other members of your class.

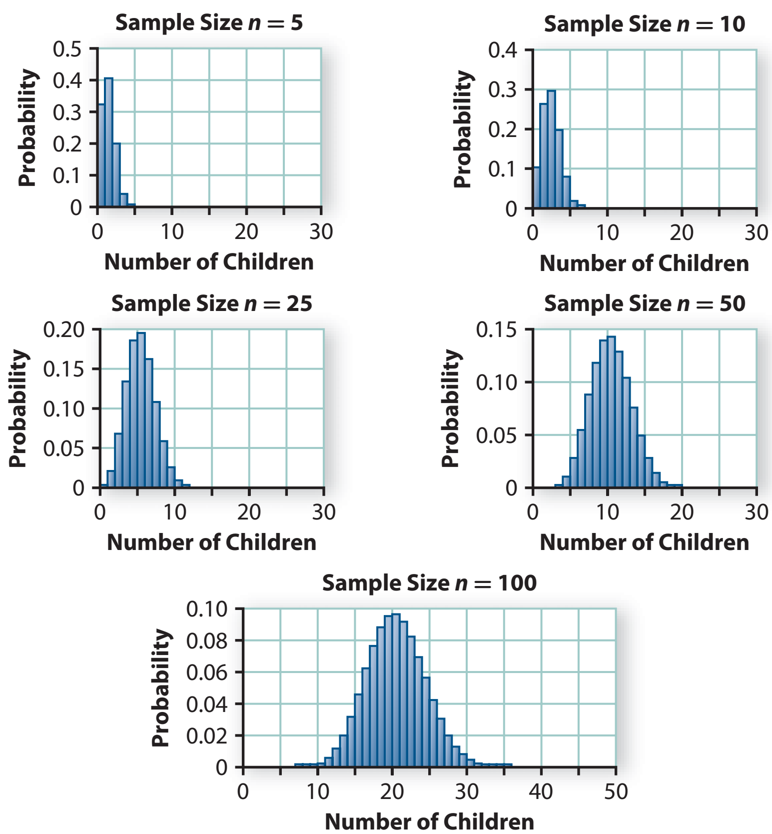
d. Perform 200 runs using the software or combine calculator results with the rest of your class until you have the results from 200 runs. Make a histogram of this **approximate binomial distribution**. What is its shape?

e. Estimate the expected number of heads and the standard deviation from the histogram. How does the expected number compare to your answer from Part a?

CPMP-Tools

In Problem 4, you saw a binomial distribution that was approximately normal in shape. In the next problem, you will examine whether that is the case with all binomial distributions.

5 According to the 2000 U.S. Census, about 20% of the population of the United States are children, age 13 or younger. (Source: *Age: 2000, Census 2000 Brief*, October 2001 at www.census.gov/prod/2001pubs/c2kbr01-12.pdf) Suppose you take a randomly selected sample of people from the United States. The following graphs show the binomial distributions for the number of children in random samples varying in size from 5 to 100.



a. The histogram for a sample size of 5 has six bars, one for 0 successes, one for 1 success, one for 2 successes, one for 3 successes, one for 4 successes, and a very short bar for 5 successes.

i. Why are there more bars as the sample size increases?

ii. How many bars should there be for a sample size of n?

iii. Why can you see only 7 bars on the histogram for a sample size of 10?

b. Use the histograms on the previous page to help answer the following questions.

i. What happens to the shape of the distribution as the sample size increases?

ii. What happens to the expected number of successes as the sample size increases?

iii. What happens to the standard deviation of the number of successes as the sample size increases?

6 As you saw in Problem 5, not all binomial distributions are approximately normal. Statisticians have developed a *guideline* that you can use to decide whether a binomial distribution is approximately normal.

The shape of a binomial distribution will be approximately normal if the expected number of successes is 10 or more and the expected number of failures is 10 or more.

a. Write the guideline about when a binomial distribution can be considered approximately normal using two algebraic inequalities where n is the number of trials and p is the probability of a success.

b. In each of the following situations, decide if the sample size is large enough so that the binomial distribution will be approximately normal.

i. The five distributions in Problem 5

ii. 100 flips of a coin and counting the number of heads

iii. Rolling a die 50 times and counting the number of 3s

iv. Randomly selecting 1,200 U.S. residents in Problem 1 Part c

v. Randomly selecting 75 children in Problem 1 Part d

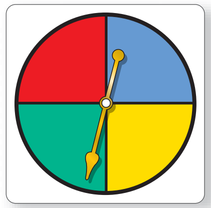
7 One of the reasons the standard deviation is such a useful measure of spread is that the standard deviation of many distributions has a simple formula. The **standard deviation of a binomial distribution** is given by:

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

a. Suppose you flip a coin 100 times and count the number of heads. Use this formula to compute the standard deviation of the binomial distribution.

b. Compare your answer in Part a to the estimate from your simulation in Problem 4 Part e.

8. Suppose you plan to spin a spinner, like the one shown, 60 times and count the number of times that you get green.



a. If p represents the probability of getting green on a single spin, what is the value of p? What does $1 - p$ represent?

b. In 60 spins, what is the expected number of times that you will get green?

c. What is the standard deviation of the number of greens?

d. Will the distribution of the number of successes be approximately normal? If so, make a sketch of this distribution, marking values of the mean and one, two, and three standard deviations from the mean on the x-axis.

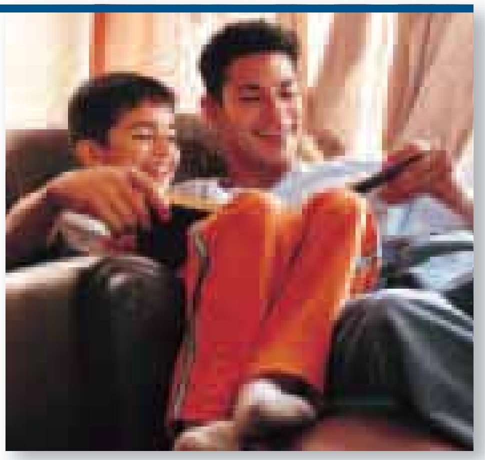
e. Would you be surprised to get 13 greens? 20 greens? Use standardized values in your explanations.

f. In Problem 5, you learned that as the number of trials for a binomial distribution increases, the standard deviation of the number of successes increases. Suppose you increase the number of spins to 240.

i. Compute the standard deviation of this distribution.

ii. Describe the relationship between the standard deviation for 240 spins and the standard deviation for 60 spins.

iii. How can you recognize this relationship by examining the formula for the standard deviation of a binomial distribution?



9. About 60% of children ages 3 to 5 are read to daily by a family member. (Source: *America's Children: Key National Indicators of Well-Being 2007*, page 51.) Suppose you take a random sample of 75 children this age.

a. Describe the shape. Compute the mean and standard deviation of the binomial distribution for this situation.

b. Using the expected number and standard deviation, describe the number of children you might get in your sample who are read to daily by a family member.

c. Out of the 75 children in your sample, suppose you get 42 children who are read to daily by a family member. Compute the standardized value for 42 children. Use this standardized value to estimate the probability of getting 42 or fewer children who are read to daily by a family member.

d. Out of the 75 children in your sample, suppose you get 50 children who are read to daily by a family member. Compute the standardized value for 50 children. Use this standardized value to estimate the probability of getting 50 or more children who are read to daily by a family member.

Summarize the Mathematics

In this investigation, you learned to use technology tools to construct a simulated binomial distribution. You also learned formulas that give the expected value and the standard deviation of a binomial distribution.

a. Describe how to use the **randBin** function of your calculator or simulation software to construct an approximate binomial distribution if you know the number of trials n and the probability of a success p.

b. Write a formula for the expected number of successes in a binomial distribution. What is another term for the expected value?

c. How can you tell whether a binomial distribution will be approximately normal in shape?

d. How do you find the spread of a binomial distribution?

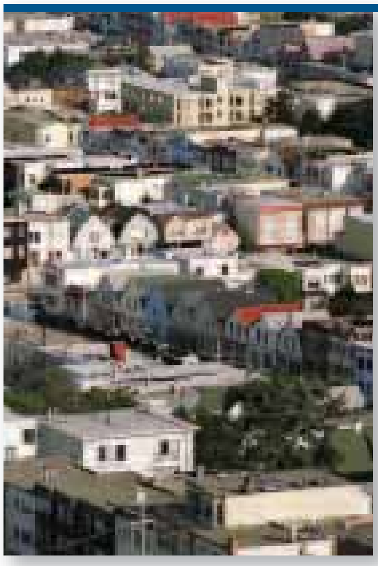
e. What happens to the expected number of successes in a binomial distribution if the probability of a success remains the same and the number of trials increases? What happens to the standard deviation?

f. How do standardized values help you analyze binomial situations?

Be prepared to share your ideas and reasoning with the class.

Check Your Understanding

About 66.2% of all housing units in the U.S. are owner-occupied. (Source: quickfacts.census.gov/qfd/states/00000.html) Suppose you randomly select 120 housing units from the U.S.



a. What is the expected number of owner-occupied units?

b. Is the binomial distribution of the number of owner-occupied units approximately normal?

c. What is the standard deviation of this binomial distribution? What does it tell you?

d. Would you be surprised to find that 65 of the housing units are owner-occupied? To find that 90 of the housing units are owner-occupied? Explain.

e. Compute a standardized value for 72 owner-occupied units. Use this standardized value to estimate the probability of getting 72 or fewer owner-occupied units.

f. Explain how to use your calculator to simulate taking a random sample of 120 housing units from the U.S. and determining how many are owner-occupied.

Investigation 2 Binomial Distributions and Making Decisions

Sometimes things happen that make you suspicious. For example, suppose your friend keeps rolling doubles in Monopoly. Knowing about binomial distributions can help you decide whether you should look a little further into this suspicious situation. As you work on problems in this investigation, look for an answer to the following question:

How do you decide whether a given probability of success is a plausible one for a given binomial situation?

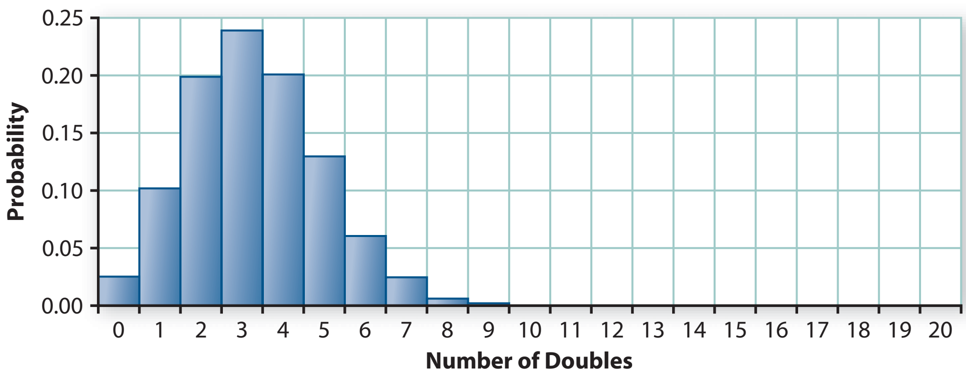
1. Suppose that while playing Monopoly, your friend rolls doubles 9 times out of 20 rolls.

   a. What is the expected number of doubles in 20 rolls?

   b. What is the standard deviation of the number of doubles? Is your friend's number of doubles more than one standard deviation above the expected number? More than two?

   c. It does seem like your friend is a little lucky. The following histogram shows the binomial distribution for $n = 20$ and $p = \frac{1}{6}$. This is the theoretical distribution with the probability of getting each number of doubles shown on the vertical axis. Should you be suspicious of the result of 9 doubles in 20 rolls?

20 Rolls of the Dice



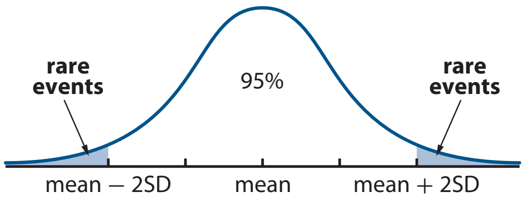
   d. What would be your reaction if your friend rolled doubles all 20 times after you started counting? Use the Multiplication Rule to find the probability that 20 doubles in a row will occur just by chance.

2 Recall that a **rare event** is one that lies in the outer 5% of a distribution. In a waiting-time distribution, rare events typically must be in the upper 5% of the distribution. In binomial distributions, rare events are those in the upper 2.5% or lower 2.5% because it is unlikely to get either an unusually large number of successes or an unusually small number of successes. Use the histogram in Problem 1 to determine which of the following are rare events if you roll a pair of dice 20 times.

a. Getting doubles only once

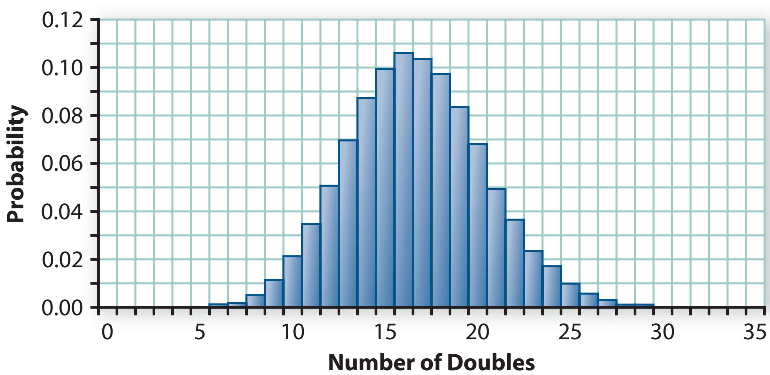
b. Getting doubles seven times

When a distribution is approximately normal, there is an easy way to identify rare events. You learned in the previous lesson that about 95% of the values in a normal distribution lie within two standard deviations of the mean. Thus, for a binomial distribution that is approximately normal, rare events are those more than two standard deviations from the mean.



3 Suppose you roll a pair of dice 100 times and count the number of times you roll doubles. The binomial distribution for this situation appears below.

100 Rolls of the Dice



a. Is this distribution approximately normal? How does it compare to the distribution in Problem 1 on page 266?

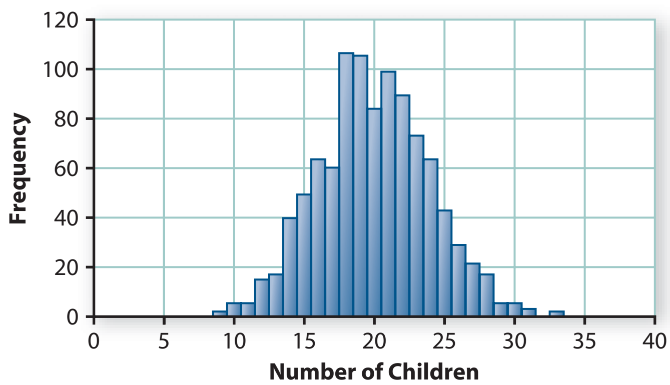
b. Look at the tails of the histogram. Approximately what numbers of doubles would be rare events?

c. Use the histogram to estimate the expected value and standard deviation of the number of doubles in 100 rolls. Then use the formulas to calculate the expected value and standard deviation.

d. Using the values you calculated in Part c, determine which numbers of doubles would be rare events in 100 rolls of the dice. Are they the same as those you identified in Part b?

4 According to the 2000 U.S. Census, about 20% of the residents of the United States are children, age 13 or younger. (Source: *Age: 2000*, *Census 2000 Brief*, October 2001.) The binomial distribution shows the number of children in 1,000 random samples. Each sample was of size $n = 100$ U.S. residents.

Number of Children in Samples of 100 U.S. Residents



a. Is this distribution approximately normal? What are the mean and standard deviation?

b. Using the mean and standard deviation you calculated in Part a, determine which numbers of children would be rare events in a random sample of 100 U.S. residents.

c. Suppose you take a randomly selected sample of 100 residents of your community and find that there are 35 children.

- **i.** Would getting 35 children be a rare event if it is true that 20% of the residents in your community are children?
- **ii.** Would this sample make you doubt that 20% is the correct percentage of residents who are children? If so, the result is called *statistically significant*.

5 Look back at the process you used to analyze the result of getting 35 children in the random sample of 100 residents in the previous problem. Review the steps that you took to help you decide whether the event of 35 children was statistically significant.

Step 1. You want to decide whether it is plausible that the given value of p is the right one for your population. State this value of p.

Step 2. For the given value of p and your sample size (number of trials), verify that the binomial distribution is approximately normal. Compute its mean and standard deviation.

Step 3. Take a random sample from your population.

Step 4. Determine how many standard deviations the number of successes in your sample is from the mean of the binomial distribution for the given p. Include a sketch of the situation.

Step 5. If the number of successes in your sample is more than two standard deviations from the mean, declare it **statistically significant**. If you have a random sample, there are two reasons that a statistically significant result could occur.

- You have the right value of p for your population but a rare event occurred.
- You have the wrong value of p for your population.

What makes the practice of statistics challenging is that it is usually impossible to know for sure which is the correct conclusion. However, it is unlikely to get a rare event (probability only 0.05) if you have the right p for your population. Thus, you conclude that the value of p is not the right one for your population.

a. How do you verify that the distribution is approximately normal? Why did you need to determine whether or not the distribution is approximately normal?

b. How do you compute the mean and standard deviation?

c. How do you find the number of standard deviations that your number of successes lies from the mean?

d. What should you conclude if you have a statistically significant result? What should you conclude if you do not have a statistically significant result?

6 About 51% of the residents of the U.S. are female. Tanner takes a random sample of 1,200 people from his community and finds that 640 of the 1,200 are female. From this, he concludes that in his community, the percentage of females is greater than 51%. Use the five steps from Problem 5 to help you decide if you agree with Tanner's conclusion.

7 In 2000, 16% of children lived in areas that did not meet one or more of the Primary National Ambient Air Quality Standards. Suppose that of a randomly selected sample of 100 children hospitalized for asthma, 29 live in an area that does not meet one or more of the Primary National Ambient Air Quality Standards. Is this a statistically significant result? What conclusion should you reach? (Source: *America's Children: Key National Indicators of Well-Being*, 2002, page 12.)

Summarize the Mathematics

In this investigation, you learned to identify rare events in binomial situations.

a How do you determine if an outcome is a rare event in a binomial distribution that is approximately normal?

b What conclusions should you consider if you read in a newspaper article that an outcome is statistically significant?

Be prepared to share your ideas and reasoning with the class.

Check Your Understanding

According to a U.S. Department of Education report (*The Condition of Education 2003*, page 43. nces.ed.gov/pubs2003/2003067_3.pdf), about 62% of high school graduates enroll in college immediately after high school graduation. Suppose you take a random sample of 200 high school graduates and count the number who enroll in college immediately after graduation.

a. What is the expected number who enroll? The expected number who do not? Is this sample size large enough that the binomial distribution for this situation will be approximately normal?

b. Complete this sentence using the mean and standard deviation.
If you select 200 high school graduates at random, you expect that ___ graduates enroll in college immediately after high school graduation, give or take ___ graduates.

c. Sketch this distribution. Include a scale on the *x*-axis that shows the mean and the points one and two standard deviations from the mean.

d. Suppose you want to see whether 62% is a plausible percentage for your community. You take a random sample of 200 high school graduates from your community and find that 135 enroll in college immediately after graduation. Is this statistically significant? What is your conclusion?

e. Suppose your random sample found 100 enrolled in college immediately after graduation. Is this statistically significant? What is your conclusion?

Applications

1 About 20% of the residents of the United States are children aged 13 and younger. Suppose you take a random sample of 200 people living in the United States and count the number of children.

a. Will the binomial distribution for the number of children in a sample of this size be approximately normal? Why or why not?

b. Compute the mean and the standard deviation of the binomial distribution for the number of children.

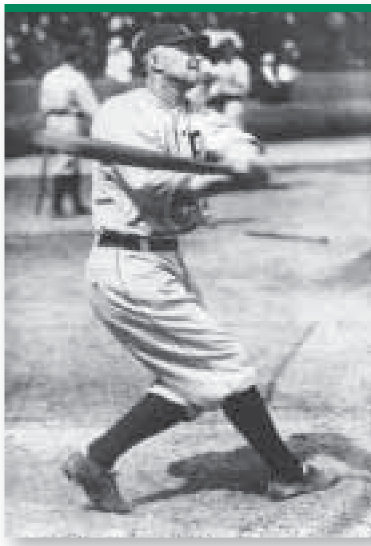
c. Make a sketch of the distribution, including a scale on the x-axis that shows the mean and one, two, and three standard deviations from the mean.

d. Use a standardized value to estimate the probability that your sample contains 45 or more children.

e. The number of children in one sample is 1.94 standard deviations below average. How many children are in the sample?

f. Suppose in your random sample of 200 people, 45 were children. Would you be surprised with that number of children? Explain.

2 Ty Cobb played Major League Baseball from 1905 to 1928. He holds the record for the highest career batting average (0.366). Suppose you pick 60 of his at bats at random and count the number of hits. (Source: baseball-almanac.com/hitting/hibavg3.shtml)



a. Is the binomial distribution of the number of hits approximately normal?

b. Compute the expected value and standard deviation of the binomial distribution of the number of hits in 60 at bats.

c. Sketch this distribution. Include a scale on the x-axis that shows the mean and the points one and two standard deviations from the mean.

d. Use a standardized value to estimate the probability that Cobb got 14 or fewer hits in this random sample of size 60.

e. Suppose you were Cobb's coach. Would you have been concerned if Cobb got only 14 hits in his next 60 at bats?

3 A library has 25 computers, which must be reserved in advance. The library assumes that people will decide independently whether or not to show up for their reserved time. It also estimates that for each person, there is an 80% chance that he or she will show up for the reserved time.

a. If the library takes 30 reservations for each time period, what is the expected number of people who will show? Is it possible that all 30 will show?

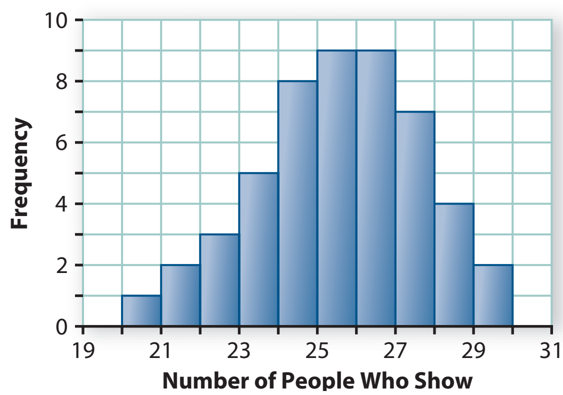
b. How many reservations should the library take so that 25 are expected to show? Is it possible that if the library takes this many reservations, more than 25 people will show?

c. Suppose the library decides to take 30 reservations. Use the **randBin** function of your calculator to conduct 5 runs of a simulation that counts the number who show. Add your results to a copy of the frequency table and histogram below so that there is a total of 55 runs.

CPMP-Tools

Computer Reservations

Number of People Who Show	Frequency (before)	Frequency (after)
19	0	
20	1	
21	2	
22	3	
23	5	
24	8	
25	9	
26	9	
27	7	
28	4	
29	2	
30	0	
Total Number of Runs	50	55



d. Describe the shape of the distribution.

e. Suppose the library takes 30 reservations. Based on the simulation, estimate the probability that more than 25 people show. Is 30 a reasonable number of reservations to take? Should the library take fewer than 30 reservations or is it reasonable to take more?

f. What assumption made in this simulation is different from the real-life situation being modeled?

4 Several years ago, a survey found that 25% of American pet owners carry pictures of their pets in their wallets. Assume this percentage is true, and you will be taking a random sample of 20 American pet owners and counting the number who carry pictures of their pets.

a. Describe how to conduct one run of a simulation using the **randBin** function of your calculator.

b. Predict the shape, mean, and standard deviation of the binomial distribution that you would get if you conduct a large number of runs.

c. Perform 10 runs of your simulation. Add your results to a copy of the following frequency table so that there is a total of 100 runs.

Number of People with Pictures of Their Pets	Frequency (before)	Frequency (after)
0	0	
1	2	
2	6	
3	12	
4	17	
5	18	
6	15	
7	10	
8	5	
9	3	
10	1	
11	1	
Total Number of Runs	90	100

d. Make a histogram of this distribution. Check your predictions in Part b.

e. Estimate the number of pet owners with pictures that would be rare events.

f. Patrick took a survey of 20 pet owners at the mall. Only one had a picture of a pet in his wallet. Does Patrick have any reason to doubt the reported figure of 25%?

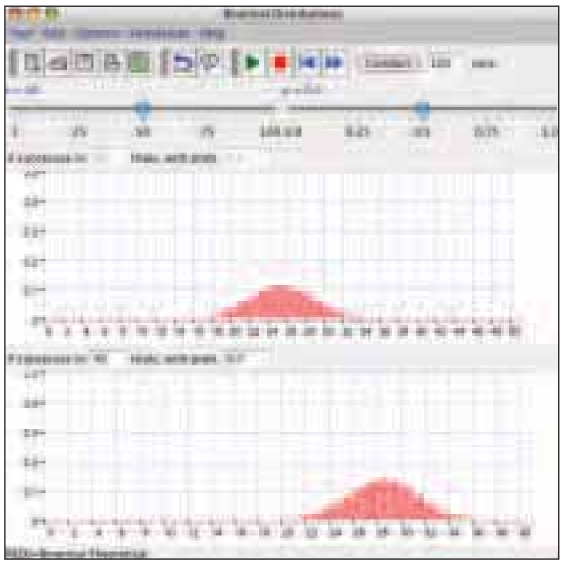
5 According to the United States census, 66.2% of housing units are occupied by the owners. (Source: quickfacts.census.gov/qfd/states/00000.html) You want to see if this percentage is plausible for your county. Suppose you take a random sample of 1,100 housing units and find that 526 are owner-occupied. Is this a statistically significant result? What conclusion should you reach? Use the five steps in Problem 5 of Investigation 2 on pages 268 and 269 to help you decide.

6 You have inherited a coin from your eccentric uncle. The coin appears a bit strange itself. To test whether the coin is fair, you toss it 150 times and count the number of heads. You get 71 heads. What conclusion should you reach?

Connections

7 In Problem 5 (page 262) of Investigation 1, you examined the shape, center, and spread of binomial distributions with probability of success 0.2 and varying sample sizes. Use computer software like the "Binomial Distributions" custom tool to help generalize your findings from that problem (Part a) and to explore a related question (Part b).

CPMP-Tools



a. As n increases, but the probability of a success p remains the same, what happens to the shape, center, and spread of the binomial distribution for the number of successes?

b. As p increases from 0.01 to 0.99, but n remains the same (for example, $n = 50$), what happens to the shape, center, and spread of the binomial distribution for the number of successes?

8 Match each description of a binomial distribution in Parts a–d with its corresponding graph. It will help if you count the number of visible bars and if you compute the standard deviations from the values in Parts a–d.

a. $p = 0.1$, $n = 40$

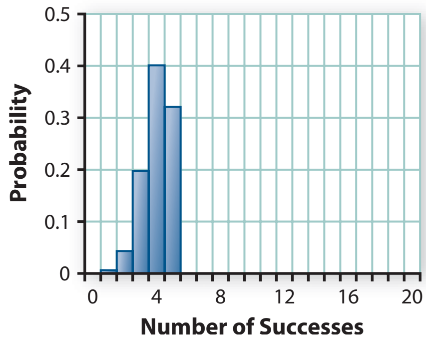
b. $p = 0.2$, $n = 20$

c. $p = 0.4$, $n = 10$

d. $p = 0.8$, $n = 5$

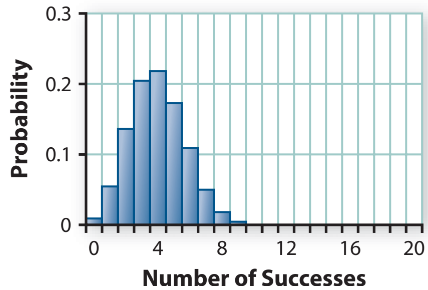
Graph I

Graph II



Graph III

Graph IV



9 In Investigation 1, you examined a simple formula for the standard deviation of a binomial distribution of the number of successes with n trials and probability of success p. In this task, you will further analyze the formula to make sure it gives the values you would expect.

a. What do you think should be the standard deviation of the binomial distribution when $p = 1$? Does the formula give this value?

b. What do you think should be the standard deviation of the binomial distribution when $p = 0$? Does the formula give this value?

c. If you know the standard deviation of the binomial distribution with n trials and probability of success p, what is the standard deviation of the binomial distribution with n trials and probability of success $1 - p$? Why does this make sense?

d. For a fixed number of trials n, what value of p will make the standard deviation, $\sqrt{np(1 - p)}$, the largest?

10 In Problem 6 of Investigation 1, you found that the sample size of 50 for rolling a die and counting the number of 3s was not large enough so that the binomial distribution can be considered approximately normal. What is the smallest sample size needed to be able to consider the binomial distribution for the situation of rolling a die and counting the number of 3s approximately normal?

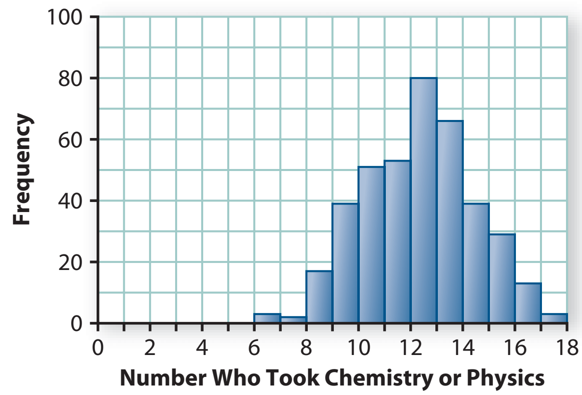
11 About 60% of high school transcripts in the United States show that the student has taken a chemistry or physics course.

a. Describe how to use the **randBin** function of your calculator to conduct a simulation to estimate the number of students who took a high school chemistry or physics course in a randomly selected group of 20 high school graduates.

b. Conduct five runs of your simulation. Add your results to a copy of the frequency table and histogram below so there is a total of 400 runs.

Chemistry and Physics on Student Transcripts

Number Who Took Chemistry or Physics	Frequency (before)	Frequency (after)
6	3	
7	2	
8	17	
9	39	
10	51	
11	53	
12	80	
13	66	
14	39	
15	29	
16	13	
17	3	
Total Number of Runs	395	400



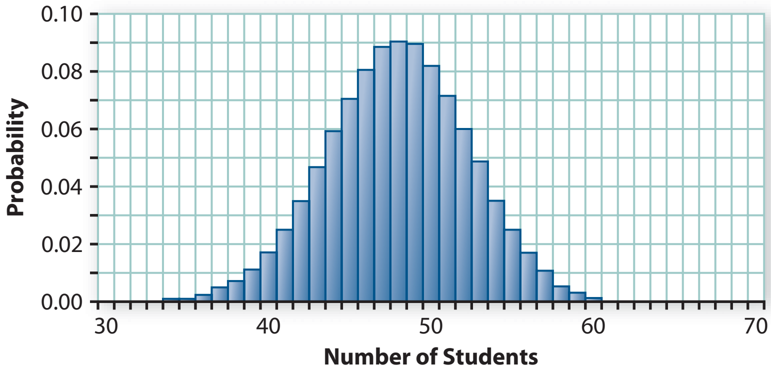
c. When 20 students are randomly selected, what is your estimate of the probability that fewer than half took chemistry or physics in high school?

d. Describe the shape of the distribution. According to the rule, should it be approximately normal?

e. Estimate the mean and standard deviation of this distribution from the histogram. Check your estimates using the formulas for the mean and standard deviation.

f. The histogram below shows the binomial distribution for the number of transcripts with chemistry or physics when 80 transcripts are selected at random, rather than 20. Describe how the shape, center, and spread of this histogram are different from those for sample sizes of 20.

Chemistry or Physics Enrollment



g. Is it more likely to find that fewer than half of the high school transcripts show that the student has taken a chemistry or physics course if you take a sample of size 20 or a sample of size 80? Explain why this should be the case.

Reflections

12 The following situations are not binomial situations. For each situation, list which of the four characteristics of a binomial situation on page 260 do not hold.

a. Roll a die 35 times. On each roll, note the number that appears on top.

b. Select five students from your classroom. Note whether each is male or female. The method of selection is to write the name of each student in your classroom on a slip of paper. Select a slip at random. Do not replace that slip. Select another slip at random. Do not replace that slip either. Select another slip at random. Continue until you have five names.

c. Place four new batteries in a calculator. Count the number that need to be replaced during 200 hours of use.

d. Flip a coin. Count the number of trials until you get a head.

e. Select 25 cars at random. Count the number of cars of each color.

13 Look back at the formulas for calculating the mean and standard deviation of a binomial probability distribution (pages 261 and 263). Consider a binomial situation with probability p.

a. If you double the sample size, what happens to the mean? To the standard deviation?

b. How does the mean vary with the sample size? How does the standard deviation vary with the sample size?

c. How are your answers to Parts a and b seen in the distributions on page 262?

14 Sports writers and commentators often speak of *odds* rather than probabilities. When you roll two dice, the odds of getting 7 as the sum are 6 to 30, or 1 to 5. That is, of 36 equally likely outcomes, 6 are successes and 30 are failures. Suppose you roll a pair of dice. Give both the probability and the odds of getting each of the following events.

a. a sum of 2

b. a sum of 8

c. an even sum

15 Suppose a person starts flipping a coin and gets heads on each flip.

a. How many heads do you think it would take before you were suspicious that the person had a two-headed coin? Before you were confident the person had a two-headed coin?

b. For your two answers in Part a, compute the probability that a person who begins to flip a fair coin will get this many heads in a row just by chance.

16 Statistics is sometimes defined as learning about a population by taking a sample from that population. Some people say you can prove anything with statistics. The truth is more nearly the opposite. You cannot *prove* anything with statistics. Explain why the second statement is the better one.

Extensions

17 Suppose you want to estimate the percentage of teens who count sheep in order to fall asleep. As the Law of Large Numbers tells you, the more teens you ask, the closer your estimate should be to the true percentage that you would get if you could ask all teens. (However, that is true only if you randomly select the teens you ask rather than, for example, asking only teens you know.)

Suppose that you want to be at least 95% sure that your estimate is off by no more than 3%. You can use the formula below to estimate the number of teens you need to ask. This formula gives the sample size n needed in order to be 95% certain that your estimate is within your margin of error E.

$$n = \frac{1}{E^2}$$

In your survey, you want to estimate the percentage of teens who count sheep. You want to be at least 95% sure your estimated percentage is within $E = 0.03$ (3%) of the true percentage. Substituting 0.03 into the formula gives

$$n = \frac{1}{(0.03)^2} = \frac{1}{0.0009} \approx 1{,}111.$$

You need to ask approximately 1,111 teens.

Being "at least 95% sure" means that out of every 100 surveys you perform, you expect that in 95 or more of them, your estimated percentage is within E of the true percentage.

CPMP-Tools

a. Suppose you want to estimate the probability a loaded die lands with a 6 on top and be at least 95% sure that your estimated probability is within 5% of the actual probability. How many times should you toss the die?

b. Suppose you do a simulation to estimate a proportion. How many runs do you need to estimate the proportion to within 1%?

c. What is the margin of error associated with 200 runs of a simulation to estimate a proportion?

d. National polls often ask 1,200 randomly selected people questions such as whether they approve of the job the President is doing. What is the margin of error associated with a sample size of 1,200?

18 Use your graphing calculator or computer software to investigate the graph of the *sample size* function $n = \frac{1}{E^2}$ given in Extensions Task 17.

a. Should x or y represent the sample size? What does the other variable represent?

b. Graph this function.

c. Describe the shape of the graph of this function, including any symmetry.

d. Use the trace function to estimate the sample size needed for a margin of error of 2%. Of 3%. Of 8%.

"By a small sample we may judge the whole piece."
Miguel de Cervantes

19 **Acceptance sampling** is one method that industry uses to control the quality of the parts it uses. For example, a recording company buys blank CDs from a supplier. To ensure the quality of these CDs, the recording company examines a sample of the CDs in each shipment. The company buys the shipment only if 5% or fewer of the CDs in the sample are defective. Assume that 10% of the CDs actually are defective.

a. Suppose the recording company examines a sample of 20 CDs from each shipment.

i. Is this a binomial situation? If so, give the sample size and the probability of a "success."

ii. Design and carry out a simulation of this situation.

iii. What is your estimate of the probability that the shipment will be accepted?

b. Suppose the recording company examines a sample of 100 CDs from each shipment. Is a normal distribution a good model of the binomial distribution? If so, use it and a standardized value to estimate the probability that the shipment will be accepted.

20 Use the margin of error formula given in Extensions Task 17 to investigate the sample size needed to cut the margin of error in half.

a. What is the margin of error for a sample size of 100? What sample size would you need to cut this margin of error in half?

b. What is the margin of error with a sample size of 625? What sample size would you need to cut this margin of error in half?

c. In general, to cut a margin of error in half, how must the sample size change? Use the formula to justify your answer.

21 The Martingale is an old gambling system. At first glance, it looks like a winner. Here is how it would work for a player betting on red in roulette. The roulette wheel has 38 spaces, and 18 of these spaces are red. On the first spin of the wheel, Mr. Garcia bets $1. If red appears, he collects $2 and leaves. If he loses, he bets $2 on red on the second spin of the wheel. If red appears, he collects $4 and leaves. If he loses, he bets $4 on red on the third spin of the wheel, and so on. Mr. Garcia keeps doubling his bet until he wins.

a. If Mr. Garcia plays 76 times, how many times does he expect to win?

b. If Mr. Garcia wins on his first try, how much money will he be ahead? If he wins on his third try, how much money will he be ahead? If Mr. Garcia finally wins on his tenth try, how much money will he be ahead?

c. From the gambler's point of view, what are some flaws in this system? (From the gambler's point of view, there are flaws in *every* "system.")

Review

22 Solve each of the following equations by reasoning with the symbols themselves.

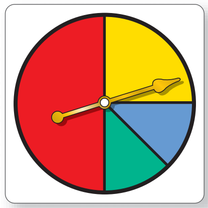
a. $6x - x^2 = 5$

b. $3(x + 9)^2 = 75$

c. $4(3^x) = 108$

d. $(x + 3)(x - 2) = 10$

23 Suppose that you spin this spinner two times.



a. What is the probability that both spins land on yellow?

b. What is the probability that the first spin lands on blue and the second spin lands on red?

c. What is the probability that one spin lands on yellow and one spin lands on green?

24 Rewrite each expression in a simpler equivalent form.

a. $4x^2 - x(3 - 2x)$

b. $6 + 3(2y - 5) - 10$

c. $3(x + 2y) + 4(3x + 2y)$

d. $x(x + 1) - (x + 5)$

25 The length of a rectangle is three times its width. The area of the rectangle is 192 cm^2. Find the perimeter of the rectangle.

26 Determine if each statement is true or false. Explain your reasoning.

a. All rectangles are similar.

b. All circles are similar.

c. If $\triangle ABC \sim \triangle DEF$ with scale factor $\frac{3}{4}$ and $\triangle ABC$ is bigger than $\triangle DEF$, then the area of $\triangle DEF$ is $\frac{3}{4}$ the area of $\triangle ABC$.

d. If $\triangle XYZ$ is a right triangle with $m\angle Y = 90°$, then $\cos X = \sin Z$.

27 What is the difference between independent events and mutually exclusive events?

On Your Own

CPMP-Tools

28 Without using a calculator or computer graphing tool, match each equation with the correct graph. Then use technology to check your answers. All graphs use the same scales.

a. $y = \frac{1}{x^2}$

b. $y = x^2 + 1$

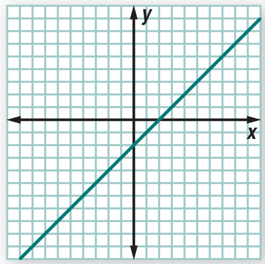
c. $y = -x + 1$

d. $y = \frac{1}{x}$

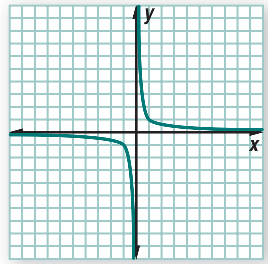
e. $y = 2^x$

f. $y = x - 1$

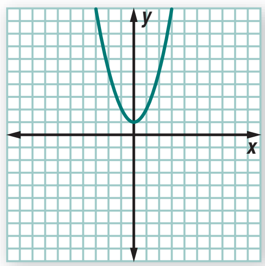
I



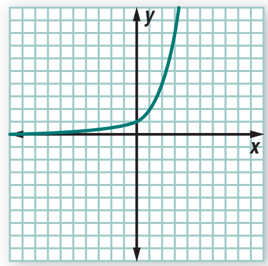
II



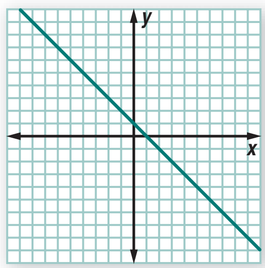
III



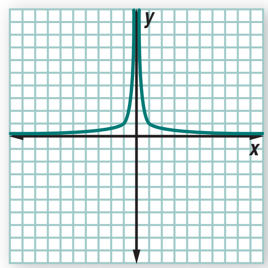
IV



V



VI



LESSON 3

Statistical Process Control

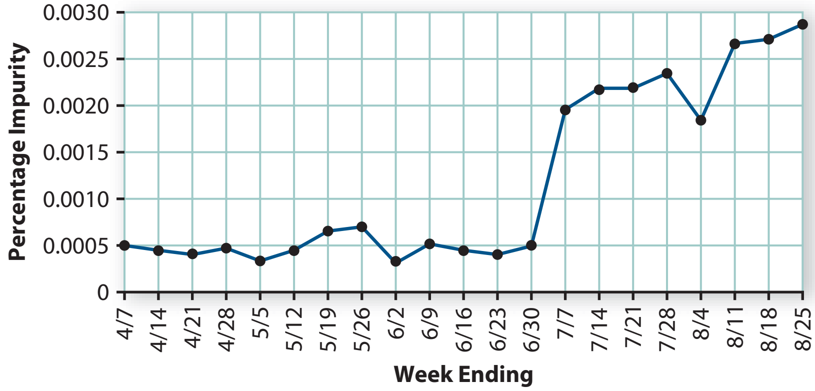
A major West Coast metal producer received the following complaint from a customer. The customer said that the metal he had recently received had an impurity in it. (The impurity was a trace element in the metal that affected how the metal performed.) Since this particular impurity had not been a problem in the past, the metal producer had not been monitoring it. The metal producer looked up the records on metal recently shipped to the customer. The percentage of the impurity in the metal for each week's shipment is given below in the table.

Metal Impurity

Week Ending	Percentage	Week Ending	Percentage	Week Ending	Percentage
4/7	0.000533	5/26	0.000721	7/14	0.002192
4/14	0.000472	6/2	0.000331	7/21	0.002205
4/21	0.000426	6/9	0.000537	7/28	0.002372
4/28	0.000481	6/16	0.000458	8/4	0.001866
5/5	0.000351	6/23	0.000420	8/11	0.002691
5/12	0.000471	6/30	0.000500	8/18	0.002721
5/19	0.000661	7/7	0.001976	8/25	0.002887

Source: From "Metal Impurity" by Lynda Finn. Copyright © Oriel Incorporated (formerly Joiner Associates), 1993. All rights reserved. Reprinted with permission.

The metal producer graphed the data on a **plot over time**. This is called a **run chart** when used for an industrial process. The run chart is shown below.



The metal producer checked with its supplier and found that the supplier had substituted a different raw material, which contained the impurity, in place of its regular raw material. The metal producer now routinely plots impurity levels of all types of impurities so that the customer will never again be the first to know.

Think About This Situation

Think about this example of variability in industrial production.

a When do you think the metal producer first started using the different raw material?

b The percentage of impurity varies even for raw material that comes from the same source. How could you estimate the variation in the percentage of impurity in shipments from the same source?

c Companies want to stop production as soon as possible after a process goes **out of control**. In the case of the metal producer, it was obvious when that happened. Often, a process goes out of control much more gradually. What should a company look for in a run chart to indicate its process might have gone out of control? Try to list several tests or rules that would signal that the process may have gone out of control.

In this lesson, you will explore some of the methods that industry uses to identify processes that have gone out of control.

Investigation 1 Out of Control Signals

In this investigation, you will examine what the run chart of an out-of-control process can look like. You will then explore some of the tests employed by industry to signal that a process may have gone out of control. Think about answers to the following question as you work through this investigation:

What does a run chart look like when the process has gone out of control?

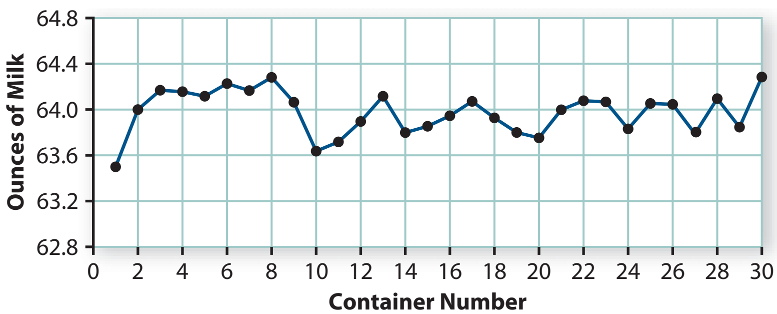
1. At one factory, a process of filling milk containers is supposed to give a distribution of weights with mean 64 ounces and standard deviation 0.2 ounces. The following four run charts come from four different machines. For each machine, 30 milk containers were filled and the number of ounces of milk measured. Two of the four machines are out of control.

   a. Which two machines appear to be out of control?

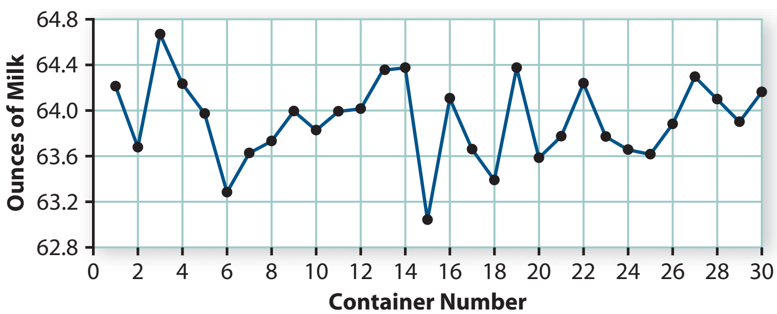
   b. On which of these machines did the mean change?

   c. On which of these machines did the standard deviation change?

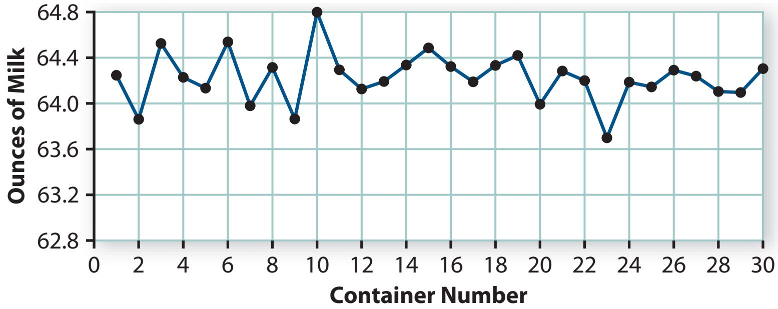
Machine 1



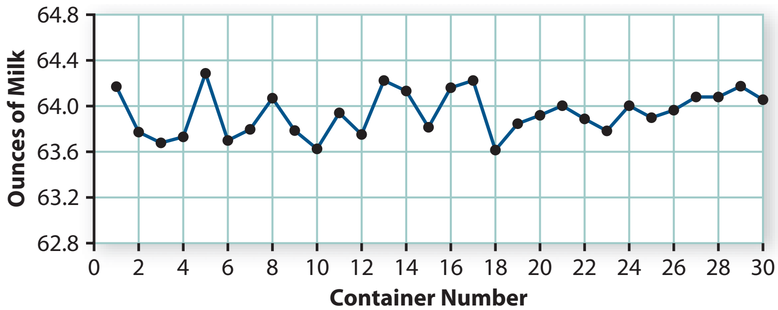
Machine 2



Machine 3



Machine 4

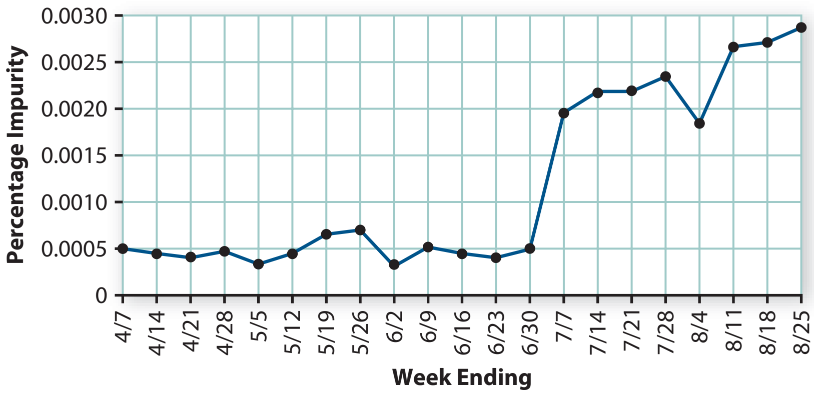


In Problem 1, you saw two ways that an out-of-control machine might behave. Since the people who monitor machines want to stop a machine as soon as possible after it has gone out of control, they have signs and patterns they look for in run charts.

2 Re-examine the data from the West Coast metal producer, reproduced below.

Metal Impurity

Week Ending	Percentage	Week Ending	Percentage	Week Ending	Percentage
4/7	0.000533	5/26	0.000721	7/14	0.002192
4/14	0.000472	6/2	0.000331	7/21	0.002205
4/21	0.000426	6/9	0.000537	7/28	0.002372
4/28	0.000481	6/16	0.000458	8/4	0.001866
5/5	0.000351	6/23	0.000420	8/11	0.002691
5/12	0.000471	6/30	0.000500	8/18	0.002721
5/19	0.000661	7/7	0.001976	8/25	0.002887



a. In the first 13 weeks of metal production, the percentage of the impurity was under control. That is, the percentage of the impurity varied a little from week to week but was of a level acceptable to the customer. Make a plot that displays the variability in these 13 percentages. Estimate the mean and standard deviation. Then compute the mean and standard deviation and compare to your estimates.

b. Now look at the percentage of impurity for the 14th week, which ended July 7. How many standard deviations from the mean computed in Part a is this percentage?

c. Assume that when the level of impurity is under control, the percentages of impurity are normally distributed with the mean and standard deviation that you calculated in Part a. If the level of impurity is under control, what is the probability of getting a percentage as high or higher than the one for July 7?

3 One commonly used test declares a process out of control when a single value is more than three standard deviations from the mean. This test assumes that the individual values are approximately normally distributed. If the process is in control, what is the probability that the next measurement will be more than three standard deviations from the mean?

4 Each of the run charts below was made from a process that was supposed to be normally distributed with a mean of 5 and a standard deviation of 1. The charts are based on displays produced using the statistical software Minitab.

a. UCL means "upper control limit." LCL means "lower control limit." How were these limits computed?

b. Using the test of three standard deviations or more from the mean, on which of the three charts is there a point where the process should be suspected to be out of control?

Chart 1

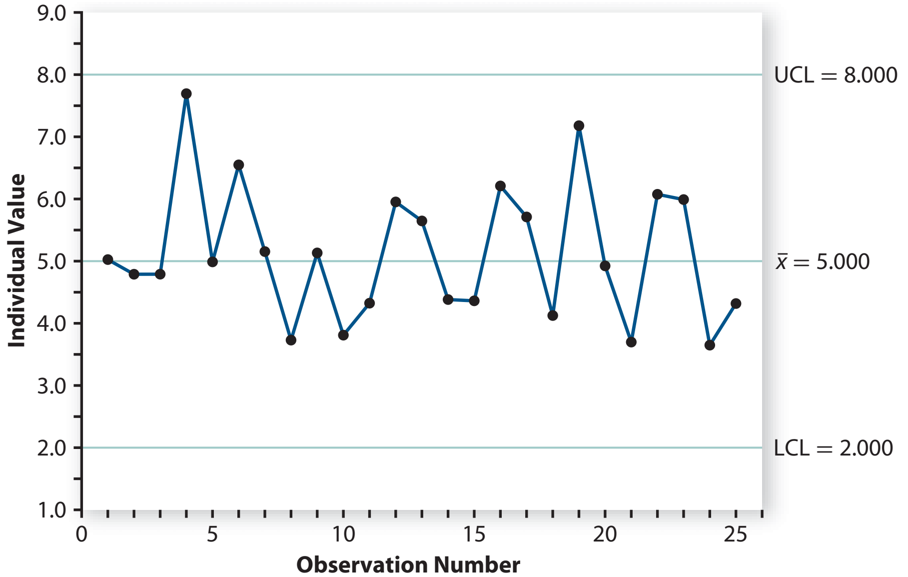


Chart 2

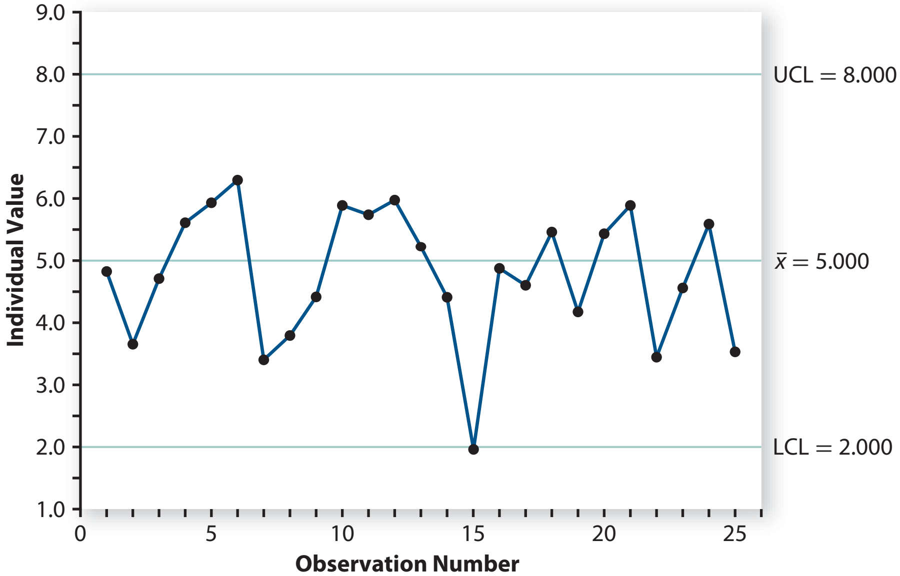
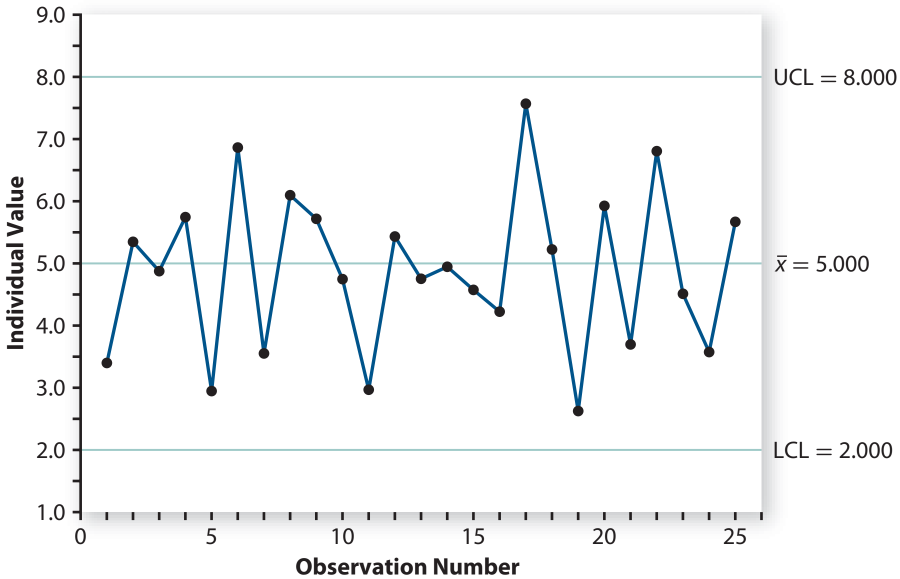
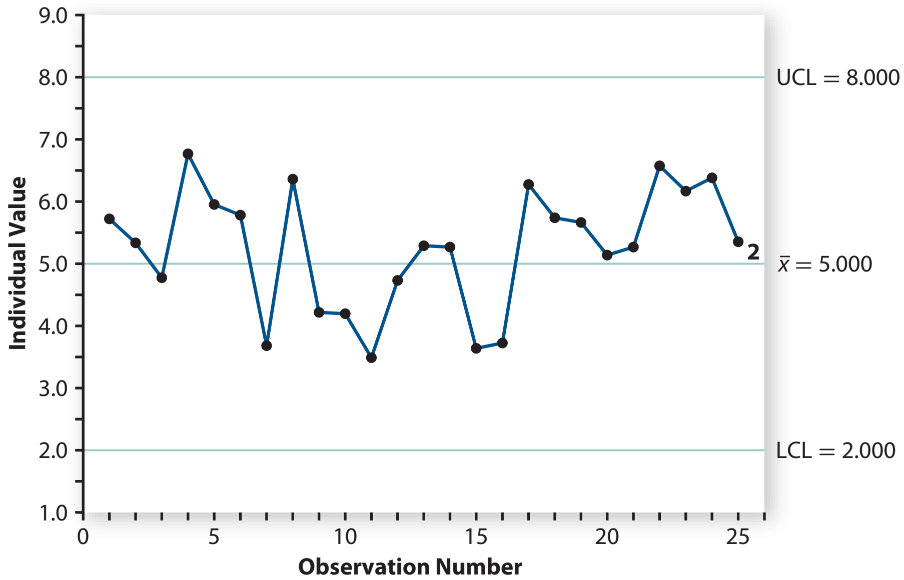


Chart 3



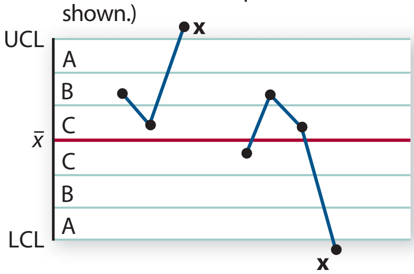
5 The process documented on the run chart below is supposed to be normally distributed with a mean of 5 and a standard deviation of 1. The small "2" below the final point, where the process was stopped, indicates that the process may have gone out of control. Why do you think Minitab has declared the process out of control?



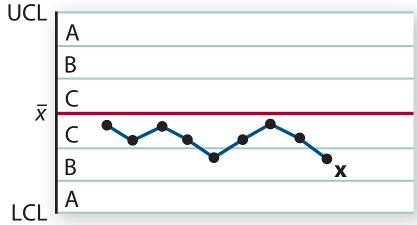
On the chart "Illustrations of Tests," which is reproduced on the next page, there are eight tests used by industry to signal that a process may have changed. The zones are marked off in standard deviations. For example, if a value falls in Zone A, it is more than two but less than three standard deviations from the mean $\bar{x}$. Each **x** marks the value at which a process is first declared out of control. Each of these tests assumes that the individual values come from a normal distribution.

Illustrations of Tests

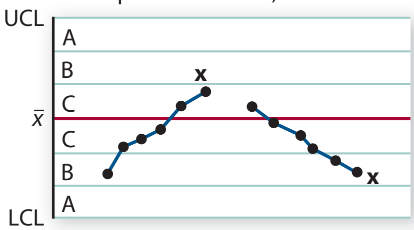
Test 1. One observation beyond either Zone A. (Two examples are shown.)



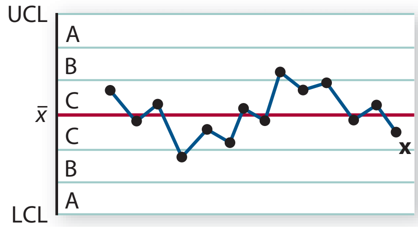
Test 2. Nine observations in a row on one half of the chart.



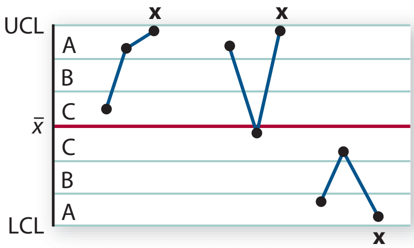
Test 3. Six observations in a row steadily increasing or decreasing. (Two examples are shown.)



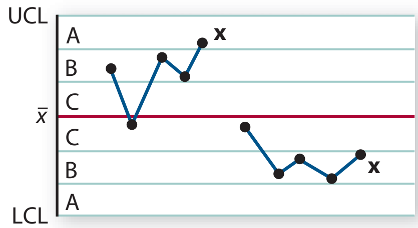
Test 4. Fourteen observations in a row alternating up or down.



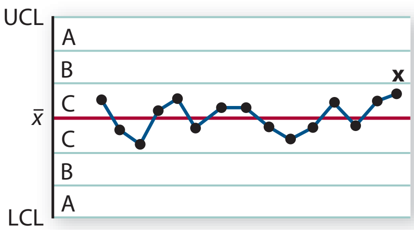
Test 5. Two out of three observations in a row on one half of the chart and in Zone A or beyond. (Three examples are shown.)



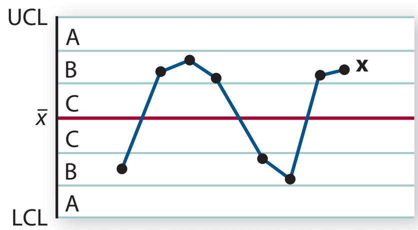
Test 6. Four out of five observations in a row on one half of the chart and in Zone B or beyond. (Two examples are shown.)



Test 7. Fifteen observations in a row within the two C zones.



Test 8. Eight observations in a row with none in either Zone C.



Source: *Western Electric. Statistical Quality Control Handbook.* Chicago: American Telephone and Telegraph Company, 1956.

6 Match each of the eight tests to the best description.

a. The observations are gradually getting larger (or smaller).

b. One observation is very far from the mean.

c. Two of three observations are unusually high (or low).

d. Four of five observations are all somewhat high (or low).

e. The mean seems to have decreased (or increased).

f. The standard deviation seems to have decreased.

g. The standard deviation seems to have increased.

h. The process shows a non-random pattern that should be explained.

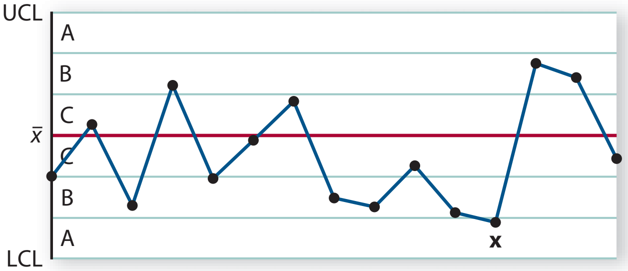
7 Look back at how these tests relate to your previous work in this investigation.

a. What is the number of the test that signals when a single value is more than three standard deviations from the mean?

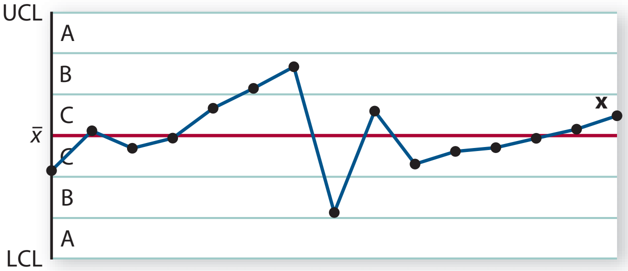
b. Which test signaled the point marked "2" in Problem 5 (page 289)?

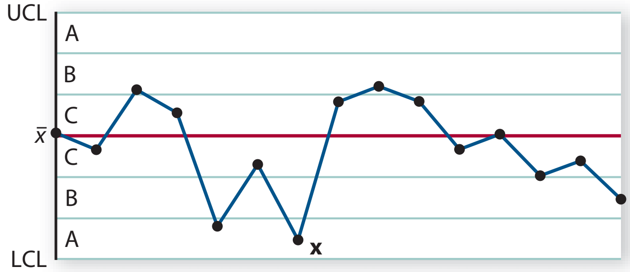
8 For each of the run charts below and on the next page, there is an **x** at the point when the process first was declared to be out of control. Give the number of the test used to decide that the process was out of control.

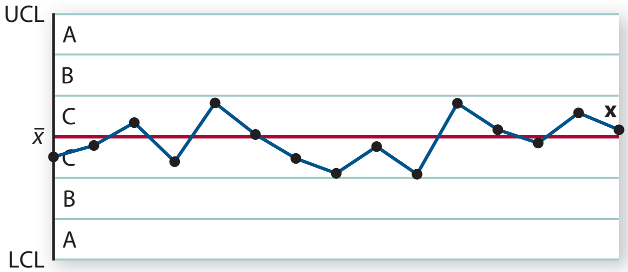
a.



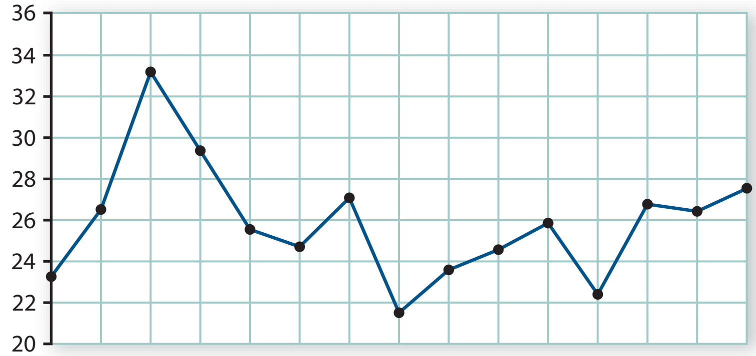
b.



c. 

d. 

9 Here is a run chart for a process that is supposed to have a normal distribution with mean 28 and standard deviation 2.



a. On a copy of the chart, identify and label the horizontal lines dividing the zones.

b. When did the process first go out of control? By which test?

c. Has the mean or the standard deviation changed?

The run charts that you have been examining are sometimes called Shewhart control charts. Dr. Walter A. Shewhart invented these charts in 1924 and developed their use while he worked for the Western Electric Company and Bell Laboratories. These charts provide a quick, visual check if a process has changed or gone "out of control." When there is change in an industrial process, the machine operator wants to know why and may have to adjust the machine.

Dr. Walter A. Shewhart

Summarize the Mathematics

In this investigation, you have examined eight tests used by industry to signal that a process may have gone out of control.

a. Describe each of these tests in your own words.

b. Name three tests that will detect a change in the mean of a process.

c. Which tests will detect a change in the standard deviation of a process?

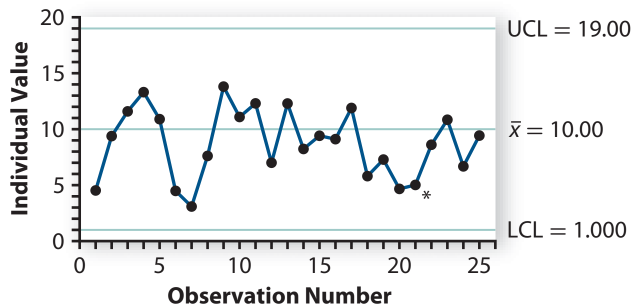
d. Suppose you are monitoring a process that is in control. What is the probability that Test 1 will signal on the very next value that the process is out of control?

Be prepared to share your ideas and reasoning with the class.

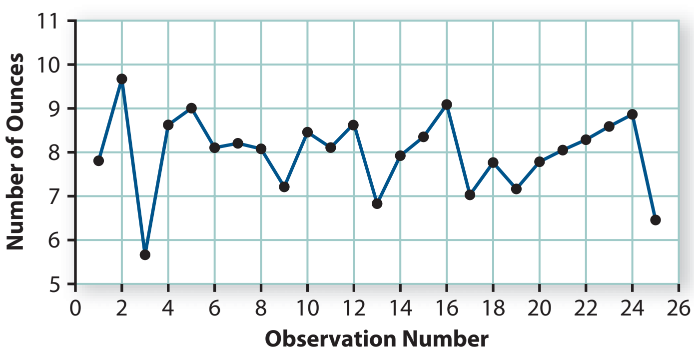
Check Your Understanding

Examine each of the following run charts.

a. For this run chart, there is an asterisk (*) below the observation at which the statistical software warned that the process may have gone out of control. Give the number of the test used.



b. The process graphed on the run chart below is supposed to have a normal distribution with mean 8 ounces and standard deviation 1 ounce. Find the point at which the process should first be declared out of control. Give the number of the test used.



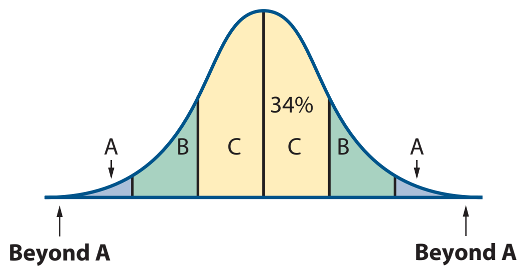
Investigation 2 False Alarms

Even a process that is under control exhibits variation. Consequently, the eight tests on page 290 occasionally will give an out-of-control signal, called a **false alarm**, for a process that is under control. For example, if you watch a process that is in control long enough, eventually one observation will be beyond one of the A zones, or six observations in a row will be steadily increasing, and so on. The tests have been designed so that false alarms occur very rarely. If you have been monitoring a process that is under control, the probability of a false alarm on the next set of observations is very small.

As you work through this investigation, look for answers to this question:

How can you find the probability of getting a false alarm?

1 Assuming that the observations of a process are normally distributed, fill in the "Percentage of Observations in Zone" on a copy of the chart below. You can then refer to the chart when working on the remaining problems.



	Zone	Number of Standard Deviations from Mean	Percentage of Observations in Zone
	Beyond Zone A		
UCL		3	
	Zone A		
		2	
	Zone B		
		1	
	Zone C		34%
Mean		0	
	Zone C		
		1	
	Zone B		
		2	
	Zone A		
LCL		3	
	Beyond Zone A		

2 Suppose a machine is filling cartons of ice cream and is under control. The operator uses only Test 1 to signal that the weight of the ice cream in the cartons may have gone out of control. What is the probability of a false alarm on the very next carton of ice cream being filled? Place your result in the appropriate row of a copy of the following table. Keep your table, as you will fill out some of the remaining rows in other problems.

False Alarms

Test	Probability of a False Alarm on the Next Set of Observations of a Process Under Control
1 One observation beyond either Zone A	
2 Nine observations in a row on one half of the chart	
3 Six observations in a row steadily increasing or decreasing	
4 Fourteen observations in a row alternating up and down	
5 Two out of three observations in a row on one half of the chart and in Zone A or beyond	
6 Four out of five observations in a row on one half of the chart and in Zone B or beyond	
7 Fifteen observations in a row within the two C zones	
8 Eight observations in a row with none in either Zone C	

3 Now suppose the ice cream machine operator is using only Test 7. The machine continues to stay in control.

a. What is the probability that a single observation will fall either within Zone C in the top half of the chart or within Zone C in the bottom half of the chart?

b. What is the probability that the observations from each of the next 15 cartons filled will all fall within either of the two C zones? (In this and subsequent problems, you may assume that the observations are independent, although independence of measurements is rarely attained in practice even with processes that are under control. Nevertheless, your computations will be good approximations.)

c. What is the probability that the ice cream machine operator will get a false alarm from the next 15 cartons being filled? Place your result in the appropriate row of your copy of the table, rounding to the nearest ten-thousandth.

4 Now suppose the ice cream machine operator is using only Test 2. The machine continues to stay in control.

a. What is the probability that the next nine observations are all in the bottom half of the chart?

b. What is the probability that the next nine observations are all in the top half of the chart?

c. What is the probability that the next nine observations are all in the top half of the chart or all in the bottom half of the chart?

d. What is the probability that the operator will get a false alarm from the next nine cartons being filled? Place your result in the appropriate row of your copy of the table, rounding to the nearest ten-thousandth.

5 A ceramic plate machine makes 20,000 plates in a year. It is under control. If the operator uses Test 1 only and measures every hundredth plate, what is the probability the operator will get through the year without having to stop the machine?

6 When computing the probabilities in Problems 3, 4, and 5, you used the Multiplication Rule for Independent Events. In real-life situations, the measurements will not be completely independent, even with processes that are under control. For example, when using Test 2, if the first measurement is in the bottom half of the chart, the next measurement will be slightly more likely to be in the bottom half of the chart than in the top half. Does this lack of independence mean that the probability of a false alarm when you use Test 2 in practice is greater or is less than the probability you computed in Problem 4?

The eight quality control tests were all devised so that, when you are monitoring a process that is in control, the probability of a false alarm on the next set of observations tested is about 0.005 or less. You will determine the probability of a false alarm for the remaining five tests on your chart in the On Your Own tasks.

Summarize the Mathematics

In this investigation, you examined the likelihood of a false alarm when using quality control tests.

a What is a false alarm? Why do false alarms occur occasionally if the process is under control?

b If you get an out-of-control signal, is there any way to tell for sure whether the process is out of control or whether it is a false alarm?

Be prepared to share your ideas and examples with the class.

Check Your Understanding

A machine operator is using the following rule as an out-of-control signal; six observations in a row in Zone B in the top half of the chart or six observations in a row in Zone B in the bottom half of the chart. Assume the machine is in control.

a. What is the probability that the next six values are all in Zone B in the top half of the chart?

b. What is the probability that the operator gets a false alarm from the next six values?

Investigation 3 The Central Limit Theorem

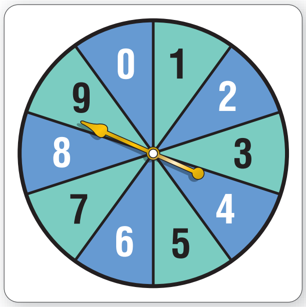
The use of control charts in the previous investigations required that the individual measurements be approximately normally distributed. However, the distribution of measurements of the parts coming off an assembly line may well be skewed. What can be done in that case?

As you work through this investigation, look for answers to the following question:

What is the Central Limit Theorem and how does it allow you to use control charts even when individual measurements come from a skewed distribution?

1. Suppose that a process is under control. How would you find the probability of a false alarm on the next measurement when using Test 1 on page 290? Where did you use the assumption that the individual values are approximately normally distributed?

2. The eight tests illustrated on page 290 assume that observations come from a distribution that is approximately normal. For individual observations, this is not always the case. For example, suppose the carnival wheel below is designed to produce random digits from the set {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}.



a. Suppose the wheel is operating correctly and produces digits at random.

i. Describe the shape of the probability distribution of all possible outcomes.

ii. What is the mean of the distribution?

iii. What is the standard deviation of the distribution?

b. Can Test 1 ever signal that this process may be out of control? Explain. What about Test 5?

c. Since the distribution in Part a is not approximately normal, you should not use the eight tests on individual digits to check whether the wheel has gone out of control. Explore what you might do so you can use the tests.

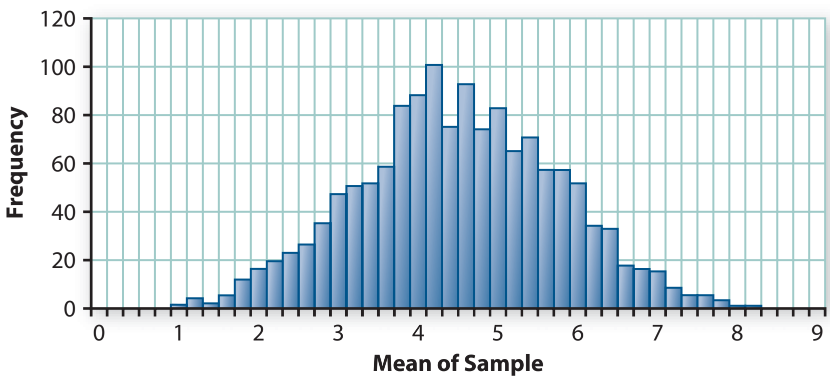
i. Use technology to produce 5 digits at random and find their mean.

ii. Nela repeated part i 1,400 times using simulation software. That is, the software selected 5 digits at random and found the mean $\overline{x}$. Then the process was repeated until there were 1,400 means. What is the smallest mean Nela could have observed? The largest?

iii. The histogram below shows Nela's 1,400 means. About how many times did she get a mean of 2? Of 4.2?

CPMP-Tools

Means of Samples of Five Random Digits



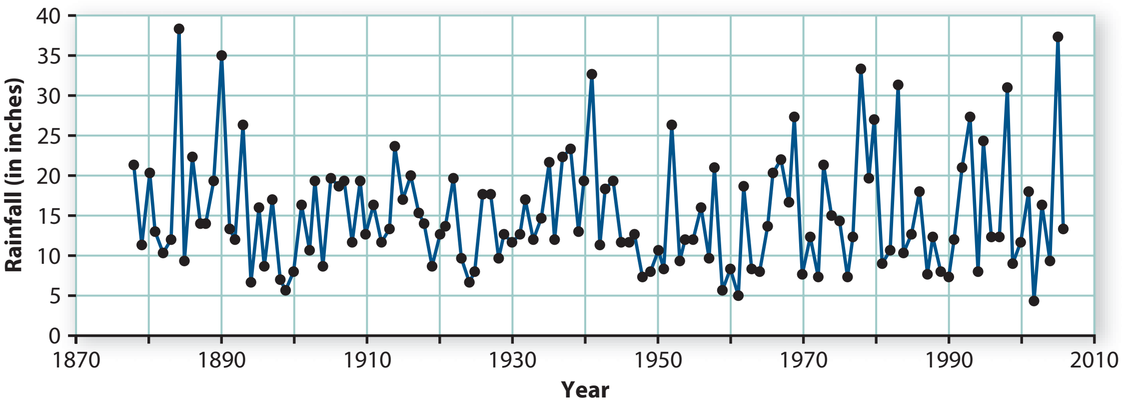
d. Use simulation software to conduct 1,500 runs of selecting 3 digits at random and calculating their means. Make a histogram of the distribution. Compare the shape and center of your histogram with Nela's histogram.

e. Repeat Part d selecting samples of 10 digits. Of 20 digits.

f. How could you use means to meet the normality assumption of the eight tests?

3 The following plot and table show the rainfall in Los Angeles for the 129 years from 1878 through 2006. The amount varies a great deal from year to year. (Source: National Weather Service)

Los Angeles Annual Rainfall 1878–2006



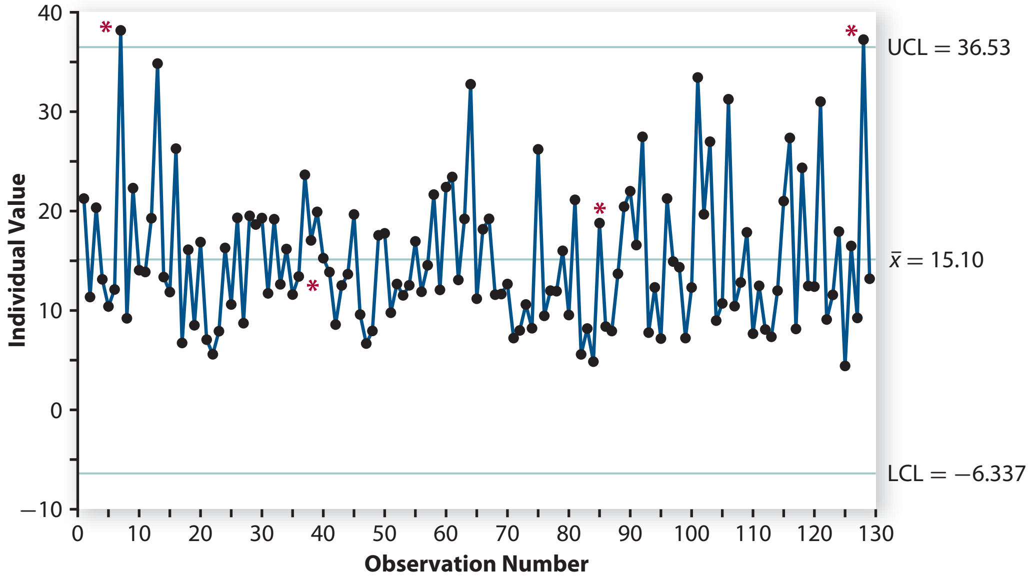
Year	Rainfall (in inches)
1878	21.26
1879	11.35
1880	20.34
1881	13.13
1882	10.40
1883	12.11
1884	38.18
1885	9.21
1886	22.31
1887	14.05
1888	13.87
1889	19.28
1890	34.84
1891	13.36
1892	11.85
1893	26.28
1894	6.73
1895	16.11
1896	8.51
1897	16.88
1898	7.06
1899	5.59
1900	7.91
1901	16.29
1902	10.60
1903	19.32
1904	8.72
1905	19.52
1906	18.65
1907	19.30
1908	11.72
1909	19.18
1910	12.63
1911	16.18
1912	11.60
1913	13.42
1914	23.65
1915	17.05
1916	19.92
1917	15.26
1918	13.86
1919	8.58
1920	12.52

Year	Rainfall (in inches)
1921	13.65
1922	19.66
1923	9.59
1924	6.67
1925	7.94
1926	17.56
1927	17.76
1928	9.77
1929	12.66
1930	11.52
1931	12.53
1932	16.95
1933	11.88
1934	14.55
1935	21.66
1936	12.07
1937	22.41
1938	23.43
1939	13.07
1940	19.21
1941	32.76
1942	11.18
1943	18.17
1944	19.22
1945	11.59
1946	11.65
1947	12.66
1948	7.22
1949	7.99
1950	10.60
1951	8.21
1952	26.21
1953	9.46
1954	11.99
1955	11.94
1956	16.00
1957	9.54
1958	21.13
1959	5.58
1960	8.18
1961	4.85
1962	18.79
1963	8.38

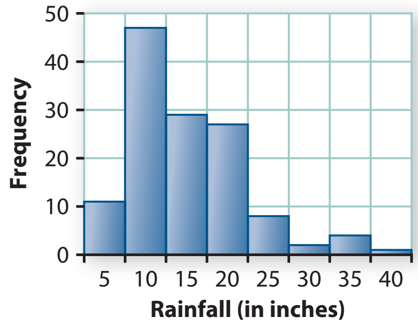
Year	Rainfall (in inches)
1964	7.93
1965	13.69
1966	20.44
1967	22.00
1968	16.58
1969	27.47
1970	7.77
1971	12.32
1972	7.17
1973	21.26
1974	14.92
1975	14.35
1976	7.22
1977	12.31
1978	33.44
1979	19.67
1980	26.98
1981	8.98
1982	10.71
1983	31.25
1984	10.43
1985	12.82
1986	17.86
1987	7.66
1988	12.48
1989	8.08
1990	7.35
1991	11.99
1992	21.00
1993	27.36
1994	8.14
1995	24.35
1996	12.46
1997	12.40
1998	31.01
1999	9.09
2000	11.57
2001	17.94
2002	4.42
2003	16.49
2004	9.25
2005	37.25
2006	13.19

a. In what year did Los Angeles receive the most rain? How many inches was this?

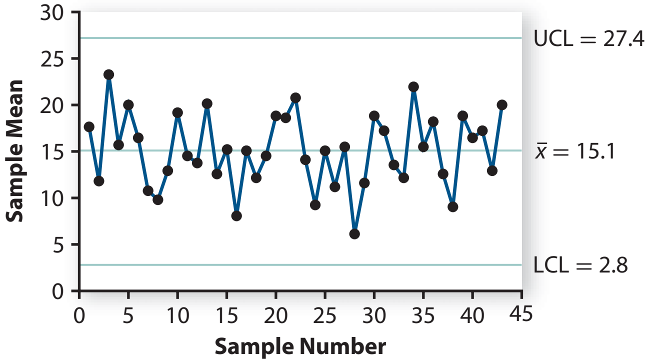
b. The control chart below was made using the historical mean yearly rainfall of 15.1 inches. According to this chart, rainfall was judged out of control at four points. Which years were these? What caused the out-of-control signals?



c. The histogram below shows the distribution of rainfall for the years 1878 through 2006. Describe its shape. Do these measurements meet the conditions for using the eight tests for deciding when a process is out of control?



d. When individual measurements have a skewed distribution, quality control practitioners can plot the *means* of samples of measurements rather than plotting individual measurements. For example, the measurements of rainfall for the first three years were 21.26, 11.35, and 20.34 inches. The mean, $\bar{x}$, for these three years was 17.65 inches. That value is the first one plotted on the ***x*-bar chart** on the following page.



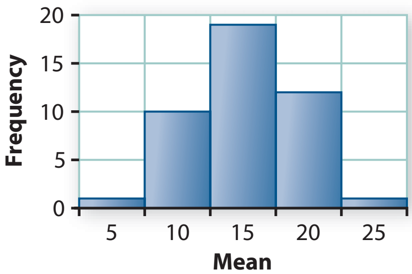
i. Using the table, find the value of the second mean plotted.

ii. Find the value of the last mean plotted.

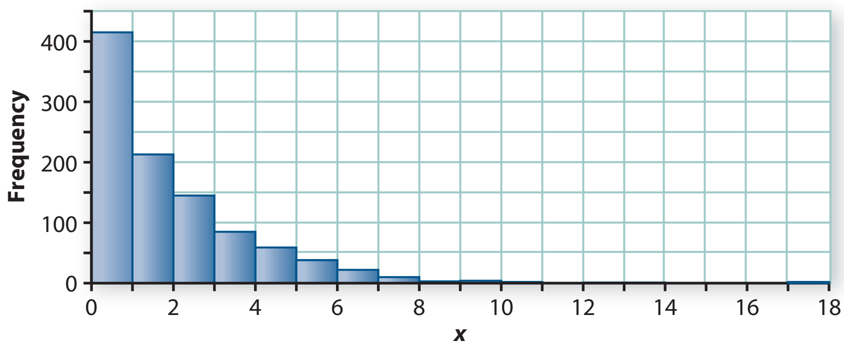
iii. Does any mean give an out of control signal?

iv. Note that the middle line of the *x*-bar chart is at 15.1, the historical mean for individual years. Explain why 15.1 is also the mean of the values on the *x*-bar chart.

e. The histogram below shows the distribution of the sample means for consecutive samples of size 3. Are conditions now met for using the eight tests?



4 The two previous problems illustrated the **Central Limit Theorem**. Imagine taking repeated random samples of a fixed size from a non-normal distribution. The larger the sample size you use, the more approximately normal the distribution of sample means $\overline{x}$ tends to be. Typically, quality control practitioners use two to five measurements per sample. In general, the more skewed the distribution of individual measurements, the larger your sample size should be. Consider the strongly right-skewed population below, which has a mean of 2.



a. Using software like the "Distribution of Sample Means" custom tool, take 100 samples of size 1 from this skewed distribution. Describe the shape. Give the mean and standard deviation of the distribution of sample means.

b. Repeat Part a for a sample size of 2. A sample size of 5. A sample size of 10. A sample size of 20.

c. Describe what happens to the mean and standard deviation of the distribution of the sample means as the sample size increases. Compare your observations with those of your classmates.

Summarize the Mathematics

In this investigation, you saw how the Central Limit Theorem can be used in situations when you want to use a control chart, but the distribution of measurements is not approximately normal.

a. What is the Central Limit Theorem?

b. How is the Central Limit Theorem used in statistical process control?

c. Describe how the shape, center, and spread of the distribution of sample means is related to sample size.

Be prepared to explain your ideas and examples with the class.

Check Your Understanding

Your company has been successful in adjusting the machine that fills cereal boxes so that the mean weight in the boxes is 16.1 ounces. After filling several hundred boxes, it is obvious that the distribution of weights is somewhat skewed. So to ensure that you are warned if the process goes out of control, you decide to use an *x*-bar chart with samples of size 3. You set the upper control limit at 16.39 and the lower control limit at 15.81. Several days later, seven samples of the weights of three individual boxes were taken, as shown in the table.

Sample	Measurements
first	16.05, 16.27, 15.68
second	16.10, 15.80, 15.75
third	16.98, 16.12, 15.73
fourth	15.93, 15.92, 15.91
fifth	16.47, 16.04, 16.93
sixth	15.58, 15.90, 16.25
seventh	15.84, 15.60, 15.77

a. Compute the mean of each sample of size 3. Plot all seven means on an *x*-bar chart that shows the upper and lower control limits.

b. Has the process gone out of control?

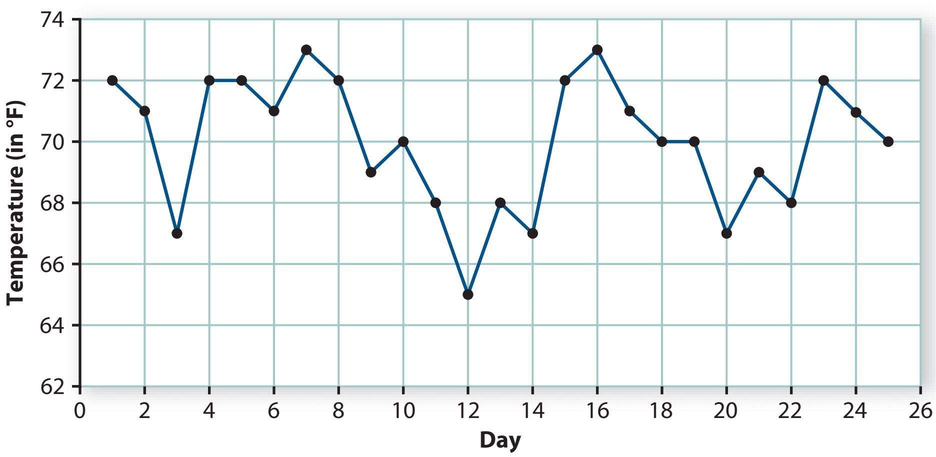
c. Why were samples of size 3 used rather than size 1?

Applications

1 The thermostat in the Simpson's apartment is set at 70° Fahrenheit. Mrs. Simpson checks the temperature in the apartment every day at noon. The table and plot below give her observations over the last 25 days. She has been satisfied with the results and felt that the process was under control. However, at noon today, the temperature in her apartment was 63°F. Should Mrs. Simpson call building maintenance? Explain why or why not, in terms of statistical process control methods.

Temperature Observations

Day	Temp (in °F)	Day	Temp (in °F)	Day	Temp (in °F)
1	72	10	70	19	70
2	71	11	68	20	67
3	67	12	65	21	69
4	72	13	68	22	68
5	72	14	67	23	72
6	71	15	72	24	71
7	73	16	73	25	70
8	72	17	71		
9	69	18	70		



2 A person is running a machine that makes nails. The nails are supposed to have a mean length of 2 inches. But, as in all processes, the nails do not come out exactly 2 inches long each time. The machine is supposed to be set so the distribution of the lengths of the nails is normal and the standard deviation of the lengths of the nails is 0.03 inches.

a. Sketch the distribution of the lengths of the nails when the machine is under control. Mark the mean and one, two, and three standard deviations from the mean on the horizontal axis.

b. If the machine is set correctly, what percentage of the nails will be more than 2.06 inches long? Less than 1.94 inches long?

c. What percentage of the nails will be more than 2.09 inches long? Less than 1.91 inches long?

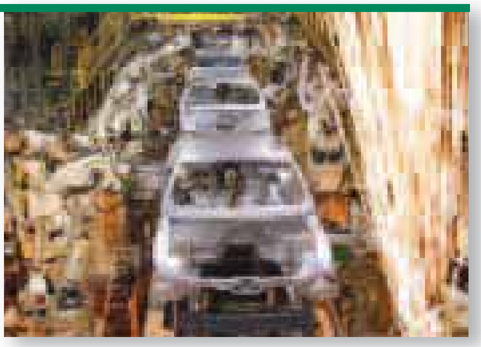
d. Suppose the machinist turns on the machine one morning after it has been cleaned. He finds these lengths in inches for the first ten nails, 2.01, 2.08, 1.97, 1.99, 1.92, 2.00, 2.03, 1.99, 1.97, and 1.95. Explain what your advice would be and why.

3 In Investigation 2, Problems 2–4, you explored the probabilities of false alarms in the context of a machine filling cartons of ice cream with the process under control. Suppose now that the operator is using only Test 8 on page 290.

a. What is the probability that the next observation is not in either Zone C?

b. What is the probability that none of the next eight observations are in either Zone C?

c. What is the probability that the operator will get a false alarm from the next eight cartons being filled? Place your result in the appropriate row of your copy of the table from Investigation 2. Round your answer to the nearest ten-thousandth.



4 The Ford Motor Company lists these four signals on its control charts.

- Any point outside of the control limits, more than three standard deviations from the mean
- Seven points in a row that are all above or all below the central line
- Seven points in a row that are either increasing or decreasing
- Any other obviously non-random pattern

Source: Ford Motor Company, *Continuing Process Control and Process Capability Improvement*. December, 1987.

a. Which of these tests is exactly the same as one of the tests (page 290) from the Western Electric handbook?

b. Assume that a manufacturing process is in control. If you are using only Ford's first test, what is the probability of a false alarm on the next observation?

c. Assume that a manufacturing process is in control. What is the probability of a false alarm on the next seven observations if you are using only Ford's second test?

d. Is Ford's second test more or less likely to produce a false alarm than the similar test from the Western Electric handbook? Explain.

e. Is Ford's third test more or less likely to produce a false alarm than the similar test from the Western Electric handbook? Explain.

5 Suppose that the ice cream machine in Investigation 2, Problem 2, is under control, and the operator is using only Test 5. You will now find the probability of a false alarm with the next three observations.

a. What is the probability that an observation will fall in Zone A at the top of the chart or beyond?

b. There are four ways that at least two of the next three observations can be in Zone A at the top of the chart or beyond. Three ways are described below. What is the fourth way?

- The first, second, and third observations are in Zone A or beyond.
- The first observation is not in Zone A or beyond. The second and third are.
- The first observation is in Zone A or beyond. The second is not, and the third is.

c. Find the probability of each of the four ways in Part b. Recall that two events are said to be **mutually exclusive** if it is impossible for both of them to occur on the same trial. Are these four ways mutually exclusive?

d. What is the probability that at least two of the next three observations will be in Zone A at the top of the chart or beyond?

e. What is the probability that at least two of the next three observations will be in Zone A at the bottom of the chart or beyond?

f. Are the events described in Parts d and e mutually exclusive?

g. What is the probability that the ice cream machine operator will get a false alarm from the next three cartons being filled? Place your result in your copy of the table from Investigation 2. Round your answer to the nearest ten-thousandth.

6 Suppose that the ice cream machine in Investigation 2, Problem 2, is under control and the operator is using only Test 6. Find the probability of a false alarm from the next five observations. Place your result in your copy of the table from Investigation 2, rounding to the nearest ten-thousandth.

7 The distribution of the lengths of widgets made by a certain process is slightly skewed. Consequently, the company uses an *x*-bar chart with samples of size 2 to determine when the process may have gone out of control. When the process is in control, the mean is 3 cm. The lower and upper control limits of the *x*-bar chart are 2.97 and 3.03. One morning, eight samples of the lengths of two individual widgets were taken, as shown in the table.

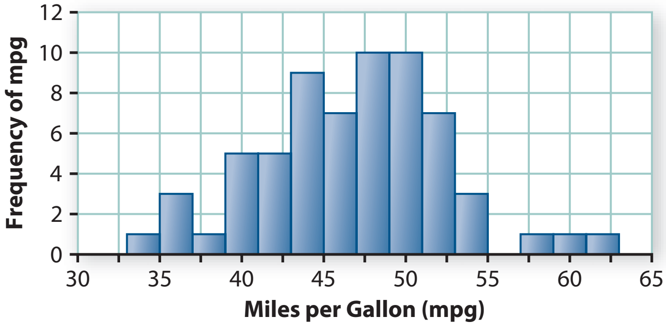
Sample	Measurements (in cm)
first	3.00, 3.02
second	3.00, 2.98
third	3.01, 2.97
fourth	3.02, 3.00
fifth	2.98, 3.00
sixth	3.00, 3.00
seventh	3.01, 3.00
eighth	2.98, 3.00

a. Compute the mean of each sample of size 2. Plot all eight means on an *x*-bar chart that shows the upper and lower control limits.

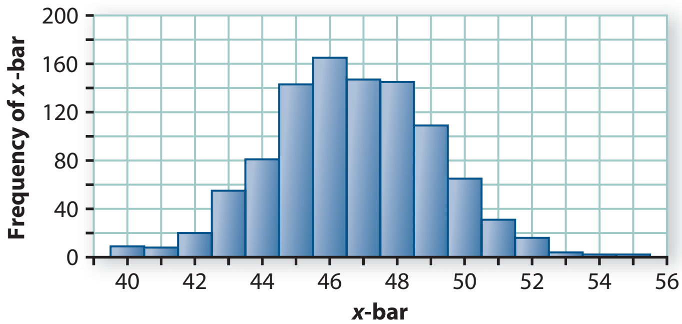
b. Has the process gone out of control?

8 At the Web site www.fueleconomy.gov, people can report their vehicles' actual gas mileage and compare it to the EPA estimate.

a. The histogram below shows the overall gas mileage reported by the first 64 owners of a 2006 Honda Civic Hybrid to make a report. The EPA estimate for the 2006 Honda Civic Hybrid was 50 miles per gallon (mpg). Does it appear that vehicles are achieving that estimate, on average?



b. The following histogram shows the means from 1,000 random samples of size 5 taken from the reported gas mileages in Part a. One of the samples had these mpg: 43.9, 53.9, 47.2, 45.0, and 46.9. Where is this sample represented on the histogram below?



c. Describe how the shape, mean, and standard deviation of the distribution of sample means differ from that of the distribution of the individual mpgs. Is this what you would expect?

Connections

9 Suppose you are operating a machine that fills cereal boxes. The boxes are supposed to contain 16 ounces. The machine fills the boxes so that the distribution of weights is normal and the standard deviation is 0.2 ounces. You can adjust the mean.

a. If you set the machine so that the mean is 16 ounces, what percentage of customers will get a box of cereal that contains less than 16 ounces?

b. Explain where you would recommend setting the mean and why.

c. Suppose you are buying a new box-filling machine. All else being equal, should you buy one with a standard deviation of 0.2 ounces or 0.4 ounces? Explain your reasoning.

10 A paper clip machine makes 4,000,000 paper clips each month. The machinist measures every 10,000th paper clip to be sure the machine is still set correctly. When the machine is set correctly, the measurements follow a normal distribution. The machinist uses only Test 1 (one value more than three standard deviations from the mean). Suppose the machine remains set correctly for a month. How many times would you expect the operator to stop production anyway?

11 Recall that two events are said to be **mutually exclusive** if it is impossible for both of them to occur on the same trial. Two events *A* and *B* are said to be **independent** if the probability that *B* occurs does not change depending on whether or not *A* occurred.

a. What is the rule for computing *P*(*A* or *B*) when *A* and *B* are mutually exclusive? For computing *P*(*A* and *B*) when *A* and *B* are independent?

b. If you pick one value at random from a normal distribution, what is the probability that it will be more than two standard deviations above the mean or more than two standard deviations below the mean?

c. If you pick one value at random from a normal distribution, what is the probability that it will be below the mean or more than two standard deviations above the mean?

d. If you pick five values at random from a normal distribution, what is the probability that all five values will be above the mean or all will be below the mean?

e. If you pick four values at random from a normal distribution, what is the probability that all four values will be more than two standard deviations above the mean?

f. If you pick five values at random from a normal distribution, what is the probability that all five values will be more than one standard deviation above the mean or all will be more than one standard deviation below the mean?

g. If you pick six values at random from a normal distribution, what is the probability that all six values will be below the mean?

h. If you select two values at random from a normal distribution, what is the probability that they both are more than two standard deviations above the mean or that they are both more than two standard deviations below the mean?

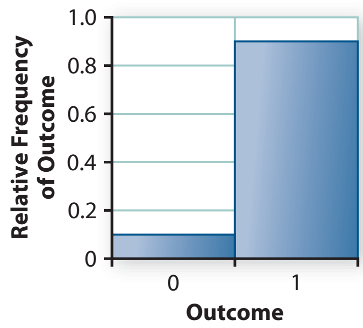
12 In the Course 2 *Probability Distributions* unit, you learned that the expected waiting time for a success in a waiting-time distribution is $\frac{1}{p}$, if p is the probability of a success on any one trial.

a. Suppose a machine operator is using only Test 1. If the process is under control, what is the expected number of items tested until Test 1 gives a false alarm? This is called the **average run length** or **ARL**.

b. If a machine operator uses both Test 1 and Test 2, is the ARL longer or shorter than if he or she uses just Test 1? Explain.

13 Here is a distribution that is *really* non-normal. Suppose that you take random samples of a fixed size n from this distribution and compute the mean. You do this several thousand times. This is exactly like making a binomial distribution except that instead of recording the number of successes in your sample, you record the number of successes divided by n, which gives the *proportion* of successes.

Outcome	Probability
0	0.1
1	0.9



a. If the sample size is 1, what do you expect the resulting distribution to look like? Describe its shape and give its mean.

b. If the sample size is 2, what is the probability you get two 0s? Two 1s? One of each? Describe the shape and give the mean of the distribution of sample means.

c. How big does the sample size have to be before the distribution of sample means is approximately normal?

Reflections

14 A machinist is making video game tokens that are supposed to be 3.2 cm in diameter.

a. What might cause the mean diameter to change?

b. What might cause the standard deviation of the diameter to change while the mean stays the same?

15 Test 7 detects a *decrease* in variability. Why might a company want to detect a decrease in variability in the manufacturing of a product or in the processing of a service?

16 Why is it best if a machine operator does not use all eight tests but picks out just a few to use?

17 If you pick six values at random from a distribution with unknown shape, what is the probability that all six values are more than the median? Can you answer this question for the mean? Explain.

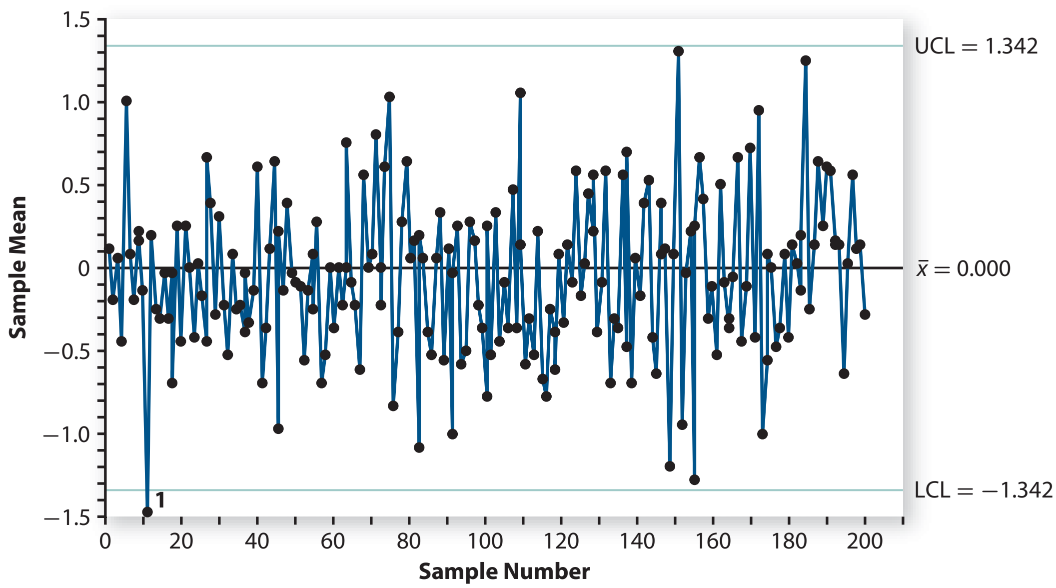
18 A false alarm occurs when a machine is in control and a test warns that it may not be.

a. What could you call a situation for which the machine is not in control and no test has given a warning?

b. The following chart has four empty cells. Two of the cells should contain the words "correct decision." Another cell should contain the words "false alarm." The fourth cell should contain your new name from Part a. Write these words in the correct cells on a copy of the chart.

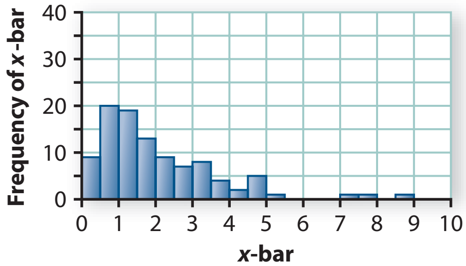
		Result of Test	
		Gives Alarm	**Does Not Give Alarm**
Condition of Machine	**In Control**		
	Not in Control		

19 The *x*-bar control chart below was made from a process that was entirely under control. Two hundred samples, each of size 5, were selected from a normal distribution with mean 0. The means of successive samples are plotted on the chart. There is one out-of-control signal, when the mean for sample 11 fell below the lower control limit. Does this mean that something is wrong or is it about what you would expect to happen?

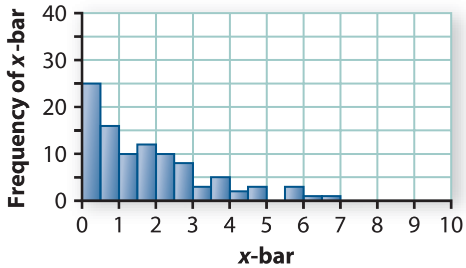


20 Look back at the distribution given in Problem 4 on page 301 of Investigation 3. Suppose 100 samples of fixed size 1, 2, 5, and 10 are taken from this population, which has mean 2. Match the sample size (1, 2, 5, or 10) with the histogram of the sample means (I, II, III, or IV).

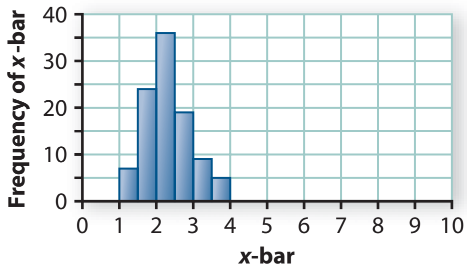
I



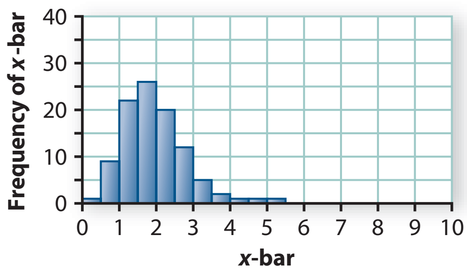
II



III



IV



Extensions

21 Suppose that the ice cream machine in Investigation 2, Problem 2, is under control and the operator is using only Test 3 from the chart on page 290. To find the probability of a false alarm, you can use the idea of permutations (ordered arrangements).

a. In how many different orders can the digits 1, 2, and 3 be listed?

b. In how many different orders can the digits 1, 2, 3, and 4 be listed?

c. Compute 4! and 3! using the *factorial* function (!) on a calculator. Compare the calculator values of 4! and 3! to your answers in Parts a and b.

d. Use the factorial function to compute the number of different orders in which the digits 1, 2, 3, 4, 5, and 6 can be listed.

e. If the digits 1, 2, 3, 4, 5, and 6 are listed in a random order, what is the probability that they are in order from smallest to largest?

f. If the digits 1, 2, 3, 4, 5, and 6 are listed in a random order, what is the probability that they are in order from largest to smallest?

g. Are the two events described in Parts e and f mutually exclusive? If any six numbers are listed at random, what is the probability that they are in order from largest to smallest or in order from smallest to largest?

h. What is the probability that the ice cream machine operator will get a false alarm using only Test 3 with the next six cartons filled? Place your result in the appropriate row of your copy of the table from Investigation 2. Round your answer to the nearest ten-thousandth.

22 The probability of a false alarm using Test 8 is much smaller than the probability of a false alarm using any of the other tests. Describe how you could change Test 8 in order to make the probability of a false alarm closer to those of the other tests.

23 Suppose that the ice cream machine in Investigation 2, Problem 2, is under control and the operator is using only Test 4. In this task, you will use simulation to estimate the probability of a false alarm with the next fourteen observations.

a. First, describe a simulation to estimate the probability that if the digits 1, 2, 3, 4, and 5 are listed in a random order. They will alternate larger, smaller, larger, smaller, larger (for example: 5, 1, 3, 2, 4). Or, they will alternate smaller, larger, smaller, larger, smaller (for example: 4, 5, 2, 3, 1).

b. Perform your simulation 20 times. What is your estimate of the probability that the digits will alternate?

c. Make a listing of all 120 possible sequences of {1, 2, 3, 4, 5}. What is the theoretical probability that if these digits are listed at random, they will alternate?

d. If you are able to share the work with others, design and carry out a simulation to estimate the probability that if the digits 1 through 14 are listed in a random order, they will alternate larger/smaller or smaller/larger. Place your result on the appropriate row of your copy of the table from Investigation 2, rounding to the nearest ten-thousandth.

24 When all possible random samples of a given size are taken from a population with standard deviation σ, the distribution of sample means has standard deviation equal to $\frac{\sigma}{\sqrt{n}}$. This is true if the samples are taken with replacement, meaning each item drawn is replaced before selecting the next one.

a. In Problem 4 of Investigation 3 (page 301) and in other problems, you noticed that the standard deviation of the distribution of sample means is smaller than the standard deviation of the population from which the random samples were taken. How are your observations explained by the above rule?

b. The population from Problem 4 has a standard deviation of about 1.93. Compare the population standard deviation to the standard deviation of the distribution of means you found for samples of size 1, 2, 5, 10, and 20.

c. The years of rainfall in Los Angeles in the table on page 299 have mean 15.1 and standard deviation 7.1. Show how the upper and lower control limits were computed in the chart in Problem 3 Part d on page 300.

Review

25 Imagine rolling a pair of tetrahedral dice.

a. Make a chart showing all possible outcomes for the sum when you roll a pair of tetrahedral dice.

b. If you roll a pair of tetrahedral dice, are the two events, getting a *sum of 5* and getting *doubles*, mutually exclusive? What is the probability that you get a sum of 5 or doubles?

26 Suppose it is true that "All graduates of Rio Grande High School must successfully complete three years of high school mathematics."

a. Write this statement in if-then form. What is the hypothesis? What is the conclusion?

b. If Alberto is a graduate of Rio Grande High School, what can you conclude? Explain your reasoning.

c. If Johanna was a student at Rio Grande High School and passed three years of high school mathematics, what can you conclude? Explain your reasoning.

27 Graph the solution set to each system of inequalities. Label the point of intersection of the boundary lines and the x- and y-intercepts of each line.

a. $x \geq 3$
$2x + y \leq 6$

b. $4x + 3y > 30$
$y < 2x - 5$

28 Without using technology tools, identify the x- and y-intercepts and the coordinates of the vertex of the graph of each quadratic function. Draw a sketch of the graph. Then check your work using technology.

a. $f(x) = (x + 3)(x - 5)$

b. $f(x) = 2x^2 - 6x$

c. $f(x) = x^2 - 4x + 3$

29 The coordinates of three vertices of $\square ABCD$ are $A(-2, 0)$, $B(1, 4)$, and $C(13, 9)$.

a. Find the coordinates of vertex D. Explain your reasoning, and verify that $ABCD$ is a parallelogram.

b. Find the measures of all the angles in the parallelogram.

c. If the diagonals of the parallelogram intersect at point E, find the length of $\overline{AE}$.

30 Delaware Valley Car Rentals has rental locations in Port Jervis, Stroudsburg, and Easton. A customer who rents a car from Delaware Valley Car Rentals can return the car to any of the three locations. Company statistics show that of the cars rented in Port Jervis, 50% are returned to Port Jervis, 20% are returned to Stroudsburg, and 30% are returned to Easton. Of the cars rented in Stroudsburg, 25% are returned to Port Jervis, 40% to Stroudsburg, and 35% to Easton. Of the cars rented in Easton, 10% are returned to Port Jervis, 30% to Stroudsburg, and 60% to Easton.

a. Represent this information in a matrix in which the rows indicate the location in which the car was rented and the columns indicate where the car was returned. Use decimal values of the percentages.

b. On one day, the company rents 20 cars in Port Jervis, 30 cars in Stroudsburg, and 45 cars in Easton. Use matrix multiplication to find the expected number of these cars that will be returned to each location.

c. Explain what the non-integer values in the matrix result might mean for this context.

31 Consider a circle with center at (−3, 1) and radius 6.

a. Draw a sketch of the circle.

b. Recall that the general form for the equation of a circle is $(x - h)^2 + (y - k)^2 = r^2$, where (h, k) is the center of the circle and r is the radius. Write the equation for this circle.

c. Calculate the area of this circle.

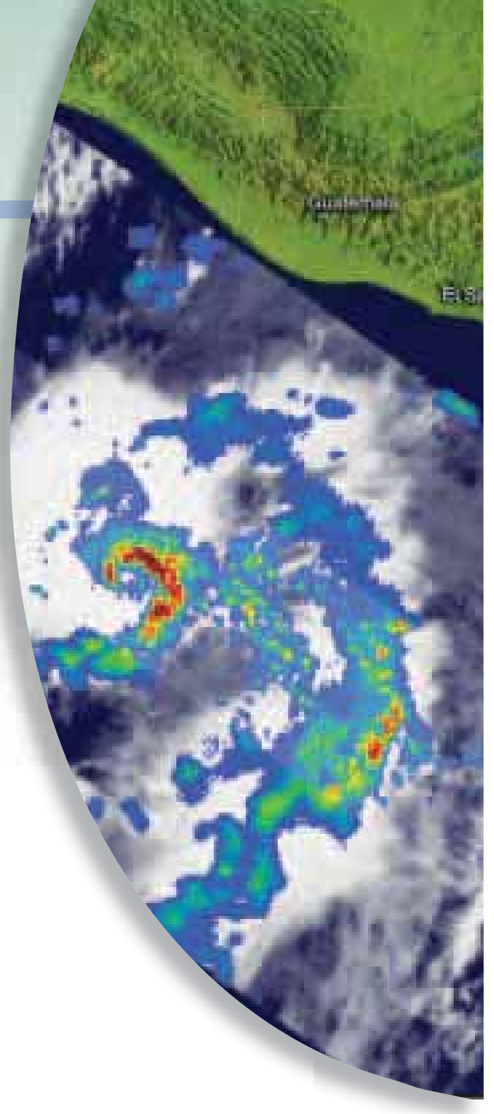
d. Calculate the circumference of this circle.

e. Write an equation for a circle with the same center but with a circumference twice the circumference of this circle.

32 If $\cos \theta = -\frac{3}{5}$ and the terminal side of θ lies in Quadrant II, find $\sin \theta$ and $\tan \theta$.

LESSON 4

Looking Back



You studied three basic tools in this unit: the normal distribution as a model of the variability in a distribution, the binomial distribution as a model of success/failure situations, and the control chart as a way to determine when an industrial process goes out of control.

The normal distribution is fundamental in characterizing variability for several reasons. As you learned in Lesson 1, many naturally occurring variables, such as human height, have an approximately normal distribution. As you learned in Lesson 2, if you have a large enough sample size, the binomial distribution is approximately normal in shape. As you learned in Lesson 3, the distribution of the means of random samples becomes more approximately normal as the sample size increases.

The following tasks will help you review and deepen your understanding of these basic ideas.

1 According to a U.S. Geological Survey report, the historical distribution of total rainfall for Guatemala averaged over 88 gauges at different stations during June, July, and August is approximately normal with a mean of 955 mm and a standard deviation of 257 mm. (Source: *June-July-August 2003 Rainfall Forecast Interpretation, Central America*, March 19, 2003)

a. Sketch this distribution. Include a scale on the horizontal axis. Mark one, two, and three standard deviations from the mean.

b. Would it be a rare event to have 500 mm of rainfall?

c. In what percentage of years is the rainfall more than 1 meter? Explain.

d. What amount of rainfall falls at the 25th percentile?

2 The claim has been made that about 57% of drivers agree that travel would be safer if the minimum driving age for new drivers was raised from 16 to 18 years old. Suppose you take a random sample of 500 drivers and ask them if they agree. (Source: www.cnn.com/2003/US/05/27/dangerous.driving/index.html)

a. If the claim is correct, what is the expected number in your sample who will agree? The expected number who will disagree?

b. Is this sample size large enough that the binomial distribution for this situation will be approximately normal?

c. Compute the mean and standard deviation of the binomial distribution.

d. Sketch this distribution. Include a scale on the horizontal axis that shows the mean and the points one and two standard deviations from the mean.

e. If the claim is true, would it be a rare event to find only 270 people in your sample who agree? Why might that happen?

3 Suppose you kept track of the gas mileage for your vehicle over a 25-week span. You recorded the data as follows.

Gas Mileage

Week	Miles per Gallon	Week	Miles per Gallon
Feb 7	23	May 9	24
Feb 14	27	May 16	27
Feb 21	27	May 23	25
Feb 28	28	May 30	28
Mar 7	25	June 6	25
Mar 14	26	June 13	26
Mar 21	25	June 20	25
Mar 28	29.5	June 27	29
Apr 4	26	July 4	26
Apr 11	27	July 11	27
Apr 18	24	July 18	24
Apr 25	26	July 25	26
May 2	26		

a. Make a run chart for your vehicle's gas mileage. What does it tell you about the consistency of your vehicle's gas mileage?

b. Find the mean and standard deviation of the measurements of miles per gallon. Draw horizontal lines on the run chart representing the mean and one and two standard deviations from the mean. Was your vehicle's mileage unusual for any week in that time period?

c. To use the eight out-of-control tests, the data must be approximately normally distributed.

i. Does that appear to be the case here? Explain.

ii. How could you use the eight tests if the distribution had not been approximately normal?

d. Would you say fuel consumption of your vehicle is in control? If not, which test is violated?

e. Five weeks after July 25, you discovered that something seemed to be wrong with your vehicle, affecting its gas mileage. Write down two different sets of mileage data for those five weeks that would indicate you had a problem. Explain the test you were using and how it would apply to your data. Use a test only once.

Summarize the Mathematics

In this unit, you learned how to use the normal distribution to model situations involving chance.

a What are the characteristics of a normal distribution?

i. What is a standardized value?

ii. How can you find the percentile of a value in a normal distribution?

b What are the characteristics of a binomial situation?

c How can you determine the shape, mean, and standard deviation of a binomial distribution?

d How can you use the mean and standard deviation of an approximately normal distribution to determine if an outcome is a rare event?

e How are the normal distribution, mean, and standard deviation used in quality control?

f What is a false alarm in quality control?

g Describe the Central Limit Theorem and how it is used in quality control.

Be prepared to share your responses with the class.

Check Your Understanding

Write, in outline form, a summary of the important mathematical concepts and methods developed in this unit. Organize your summary so that it can be used as a quick reference in future units and courses.