tell the truth about how the cards were distributed? Not necessarily. Our simulation says that it's very unlikely for someone to have to buy 23 boxes to get a full set *if* each card is equally likely to appear in a box of cereal. The evidence suggests that the company's statement is incorrect. It's still possible, however, that the NASCAR fan was just very unlucky.

CHECK YOUR UNDERSTANDING

1. Refer to the golden ticket parking lottery example. At the following month's school assembly, the two lucky winners were once again members of the AP® Statistics class. This raised suspicions about how the lottery was being conducted. How would you modify the simulation in the example to estimate the probability of getting two winners from the AP® Statistics class in back-to-back months just by chance?

2. Refer to the NASCAR and breakfast cereal example. What if the cereal company decided to make it harder to get some drivers' cards than others? For instance, suppose the chance that each card appears in a box of the cereal is Jeff Gordon, 10%; Dale Earnhardt, Jr., 30%; Tony Stewart, 20%; Danica Patrick, 25%; and Jimmie Johnson, 15%. How would you modify the simulation in the example to estimate the chance that a fan would have to buy 23 or more boxes to get the full set?

Section 5.1 Summary

- A chance process has outcomes that we cannot predict but that nonetheless have a regular distribution in very many repetitions. The **law of large numbers** says that the proportion of times that a particular outcome occurs in many repetitions will approach a single number. This long-run relative frequency of a chance outcome is its **probability**. A probability is a number between 0 (never occurs) and 1 (always occurs).

- Probabilities describe only what happens in the long run. Short runs of random phenomena like tossing coins or shooting a basketball often don't look random to us because they do not show the regularity that emerges only in very many repetitions.

- A **simulation** is an imitation of chance behavior, most often carried out with random numbers. To perform a simulation, follow the four-step process:

  STEP 4

  STATE: Ask a question of interest about some chance process.

  PLAN: Describe how to use a chance device to imitate one repetition of the process. Tell what you will record at the end of each repetition.

  DO: Perform many repetitions of the simulation.

  CONCLUDE: Use the results of your simulation to answer the question of interest.

Section 5.1 Exercises

1. **Liar, liar!** Sometimes police use a lie detector (also known as a polygraph) to help determine whether a suspect is telling the truth. A lie detector test isn't foolproof—sometimes it suggests that a person is lying when he or she is actually telling the truth (a "false positive"). Other times, the test says that the suspect is being truthful when the person is actually lying (a "false negative"). For one brand of polygraph machine, the probability of a false positive is 0.08.

(a) Interpret this probability as a long-run relative frequency.

(b) Which is a more serious error in this case: a false positive or a false negative? Justify your answer.

2. **Mammograms** Many women choose to have annual mammograms to screen for breast cancer after age 40. A mammogram isn't foolproof. Sometimes the test suggests that a woman has breast cancer when she really doesn't (a "false positive"). Other times the test says that a woman doesn't have breast cancer when she actually does (a "false negative"). Suppose the false negative rate for a mammogram is 0.10.

(a) Interpret this probability as a long-run relative frequency.

(b) Which is a more serious error in this case: a false positive or a false negative? Justify your answer.

3. **Genetics** Suppose a married man and woman both carry a gene for cystic fibrosis but don't have the disease themselves. According to the laws of genetics, the probability that their first child will develop cystic fibrosis is 0.25.

(a) Explain what this probability means.

(b) If the couple has 4 children, is one of them guaranteed to get cystic fibrosis? Explain.

4. **Texas hold 'em** In the popular Texas hold 'em variety of poker, players make their best five-card poker hand by combining the two cards they are dealt with three of five cards available to all players. You read in a book on poker that if you hold a pair (two cards of the same rank) in your hand, the probability of getting four of a kind is 88/1000.

(a) Explain what this probability means.

(b) If you play 1000 such hands, will you get four of a kind in exactly 88 of them? Explain.

5. **Spinning a quarter** With your forefinger, hold a new quarter (with a state featured on the reverse) upright, on its edge, on a hard surface. Then flick it with your other forefinger so that it spins for some time before it falls and comes to rest. Spin the coin a total of 25 times, and record the results.

(a) What's your estimate for the probability of heads? Why?

(b) Explain how you could get an even better estimate.

6. **Nickels falling over** You may feel it's obvious that the probability of a head in tossing a coin is about 1/2 because the coin has two faces. Such opinions are not always correct. Stand a nickel on edge on a hard, flat surface. Pound the surface with your hand so that the nickel falls over. Do this 25 times, and record the results.

(a) What's your estimate for the probability that the coin falls heads up? Why?

(b) Explain how you could get an even better estimate.

7. **Free throws** The figure below shows the results of a virtual basketball player shooting several free throws. Explain what this graph says about chance behavior in the short run and long run.

8. **Keep on tossing** The figure below shows the results of two different sets of 5000 coin tosses. Explain what this graph says about chance behavior in the short run and the long run.

Due for a hit A very good professional baseball player gets a hit about 35% of the time over an entire season. After the player failed to hit safely in six straight at-bats, a TV commentator said, "He is due for a hit by the law of averages." Is that right? Why?

Cold weather coming A TV weather man, predicting a colder-than-normal winter, said, "First, in looking at the past few winters, there has been a lack of really cold weather. Even though we are not supposed to use the law of averages, we are due." Do you think that "due by the law of averages" makes sense in talking about the weather? Why or why not?

11. **Playing "Pick 4"** The Pick 4 games in many state lotteries announce a four-digit winning number each day. You can think of the winning number as a four-digit group from a table of random digits. You win (or share) the jackpot if your choice matches the winning number. The winnings are divided among all players who matched the winning number. That suggests a way to get an edge.

(a) The winning number might be, for example, either 2873 or 9999. Explain why these two outcomes have exactly the same probability.

(b) If you asked many people whether 2873 or 9999 is more likely to be the randomly chosen winning number, most would favor one of them. Use the information in this section to say which one and to explain why. How might this affect the four-digit number you would choose?

12. **An unenlightened gambler**

(a) A gambler knows that red and black are equally likely to occur on each spin of a roulette wheel. He observes five consecutive reds occur and bets heavily on black at the next spin. Asked why, he explains that black is "due by the law of averages." Explain to the gambler what is wrong with this reasoning.

(b) After hearing you explain why red and black are still equally likely after five reds on the roulette wheel, the gambler moves to a poker game. He is dealt five straight red cards. He remembers what you said and assumes that the next card dealt in the same hand is equally likely to be red or black. Is the gambler right or wrong, and why?

13. **Free throws** A basketball player has probability 0.75 of making a free throw. Explain how you would use each chance device to simulate one free throw by the player.

(a) A standard deck of playing cards

(b) Table D of random digits

(c) A calculator or computer's random integer generator

14. **Stoplight** On her drive to work every day, Ilana passes through an intersection with a traffic light. The light has probability 1/3 of being green when she gets to the intersection. Explain how you would use each chance device to simulate whether the light is red or green on a given day.

(a) A six-sided die

(b) Table D of random digits

(c) A calculator or computer's random integer generator

15. **Simulation blunders** Explain what's wrong with each of the following simulation designs.

(a) A roulette wheel has 38 colored slots—18 red, 18 black, and 2 green. To simulate one spin of the wheel, let numbers 00 to 18 represent red, 19 to 37 represent black, and 38 to 40 represent green.

(b) About 10% of U.S. adults are left-handed. To simulate randomly selecting one adult at a time until you find a left-hander, use two digits. Let 00 to 09 represent being left-handed and 10 to 99 represent being right-handed. Move across a row in Table D, two digits at a time, skipping any numbers that have already appeared, until you find a number between 00 and 09. Record the number of people selected.

16. **Simulation blunders** Explain what's wrong with each of the following simulation designs.

(a) According to the Centers for Disease Control and Prevention, about 36% of U.S. adults were obese in 2012. To simulate choosing 8 adults at random and seeing how many are obese, we could use two digits. Let 00 to 35 represent obese and 36 to 99 represent not obese. Move across a row in Table D, two digits at a time, until you find 8 distinct numbers (no repeats). Record the number of obese people selected.

(b) Assume that the probability of a newborn being a boy is 0.5. To simulate choosing a random sample of 9 babies who were born at a local hospital today and observing their gender, use one digit. Use `randInt(0,9)` on your calculator to determine how many babies in the sample are male.

17. **Is this valid?** Determine whether each of the following simulation designs is valid. Justify your answer.

(a) According to a recent poll, 75% of American adults regularly recycle. To simulate choosing a random sample of 100 U.S. adults and seeing how many of them recycle, roll a 4-sided die 100 times. A result of 1, 2, or 3 means the person recycles; a 4 means that the person doesn't recycle.

(b) An archer hits the center of the target with 60% of her shots. To simulate having her shoot 10 times, use a coin. Flip the coin once for each of the 10 shots. If it lands heads, then she hits the center of the target. If the coin lands tails, she doesn't.

18. **Is this valid?** Determine whether each of the following simulation designs is valid. Justify your answer.

(a) According to a recent survey, 50% of people aged 13 and older in the United States are addicted to texting. To simulate choosing a random sample of 20 people in this population and seeing how many of them are addicted to texting, use a deck of cards. Shuffle the deck well, and then draw one card at a time. A red card means that person is addicted to texting; a black card means he isn't. Continue until you have drawn 20 cards (without replacement) for the sample.

(b) A tennis player gets 95% of his serves in play during practice (that is, the ball doesn't go out of bounds). To simulate the player hitting 5 serves, look at 5 pairs of digits going across a row in Table D. If the number is between 00 and 94, the serve is in; numbers between 95 and 99 indicate that the serve is out.

19. **Airport security** The Transportation Security Administration (TSA) is responsible for airport safety. On some flights, TSA officers randomly select passengers for an extra security check prior to boarding. One such flight had 76 passengers—12 in first class and 64 in coach class. Some passengers were surprised when none of the 10 passengers chosen for screening were seated in first class. We can use a simulation to see if this result is likely to happen by chance.

pg 296

(a) State the question of interest using the language of probability.

(b) How would you use random digits to imitate one repetition of the process? What variable would you measure?

(c) Use the line of random digits below to perform one repetition. Copy these digits onto your paper. Mark directly on or above them to show how you determined the outcomes of the chance process.

71487 09984 29077 14863 61683 47052 62224 51025

(d) In 100 repetitions of the simulation, there were 15 times when none of the 10 passengers chosen was seated in first class. What conclusion would you draw?

20. **Scrabble** In the game of Scrabble, each player begins by drawing 7 tiles from a bag containing 100 tiles. There are 42 vowels, 56 consonants, and 2 blank tiles in the bag. Cait chooses her 7 tiles and is surprised to discover that all of them are vowels. We can use a simulation to see if this result is likely to happen by chance.

(a) State the question of interest using the language of probability.

(b) How would you use random digits to imitate one repetition of the process? What variable would you measure?

(c) Use the line of random digits below to perform one repetition. Copy these digits onto your paper. Mark directly on or above them to show how you determined the outcomes of the chance process.

00694 05977 19664 65441 20903 62371 22725 53340

(d) In 1000 repetitions of the simulation, there were 2 times when all 7 tiles were vowels. What conclusion would you draw?

21. **The birthday problem** What's the probability that in a randomly selected group of 30 unrelated people, at least two have the same birthday? Let's make two assumptions to simplify the problem. First, we'll ignore the possibility of a February 29 birthday. Second, we assume that a randomly chosen person is equally likely to be born on each of the remaining 365 days of the year.

(a) How would you use random digits to imitate one repetition of the process? What variable would you measure?

(b) Use technology to perform 5 repetitions. Record the outcome of each repetition.

(c) Would you be surprised to learn that the theoretical probability is 0.71? Why or why not?

22. **Monty Hall problem** In *Parade* magazine, a reader posed the following question to Marilyn vos Savant and the "Ask Marilyn" column:

Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat. He says to you, "Do you want to pick door #2?" Is it to your advantage to switch your choice of doors?[4]

The game show in question was *Let's Make a Deal* and the host was Monty Hall. Here's the first part of Marilyn's response: "Yes; you should switch. The first door has a 1/3 chance of winning, but the second door has a 2/3 chance." Thousands of readers wrote to Marilyn to disagree with her answer. But she held her ground.

(a) Use an online *Let's Make a Deal* applet to perform at least 50 repetitions of the simulation. Record whether you stay or switch (try to do each about half the time) and the outcome of each repetition.

(b) Do you agree with Marilyn or her readers? Explain.

23. **Recycling** Do most teens recycle? To find out, an AP® Statistics class asked an SRS of 100 students at their school whether they regularly recycle. In the sample, 55 students said that they recycle. Is this convincing evidence that more than half of the students at the school would say they regularly recycle? The Fathom dotplot below shows the results of

taking 200 SRSs of 100 students from a population in which the true proportion who recycle is 0.50.
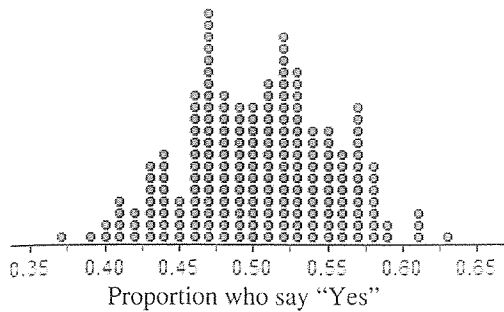
Proportion who say "Yes"

(a) Explain why the sample result does not give convincing evidence that more than half of the school's students recycle.

(b) Suppose instead that 63 students in the class's sample had said "Yes." Explain why this result would give strong evidence that a majority of the school's students recycle.

24. **Brushing teeth, wasting water?** A recent study reported that fewer than half of young adults turn off the water while brushing their teeth. Is the same true for teenagers? To find out, a group of statistics students asked an SRS of 60 students at their school if they usually brush with the water off. In the sample, 27 students said "No." The Fathom dotplot below shows the results of taking 200 SRSs of 60 students from a population in which the true proportion who brush with the water off is 0.50.

Proportion who say "No"
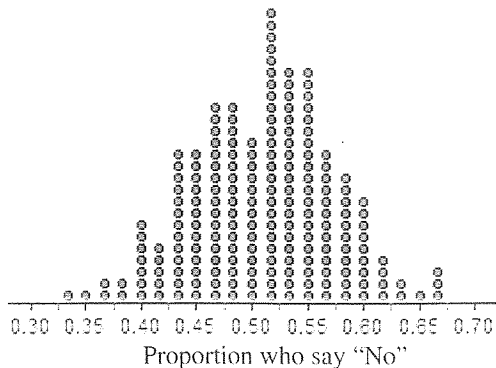
(a) Explain why the sample result does not give convincing evidence that fewer than half of the school's students brush their teeth with the water off.

(b) Suppose instead that 18 students in the class's sample had said "No." Explain why this result would give strong evidence that fewer than 50% of the school's students brush their teeth with the water off.

25. **Color-blind men** About 7% of men in the United States have some form of red-green color blindness. Suppose we randomly select 4 U.S. adult males. What's the probability that at least one of them is red-green color-blind? Design and carry out a simulation to answer this question. Follow the four-step process.

26. **Lotto** In the United Kingdom's Lotto game, a player picks six numbers from 1 to 49 for each ticket. Rosemary bought one ticket for herself. She had the lottery computer randomly select the six numbers. When the six winning numbers were drawn, Rosemary was surprised to find that none of these numbers appeared on the Lotto ticket she had bought. Should she be? Design and carry out a simulation to answer this question. Follow the four-step process.

27. **Color-blind men** Refer to Exercise 25. Suppose we randomly select one U.S. adult male at a time until we find one who is red-green color-blind. Should we be surprised if it takes us 20 or more men? Design and carry out a simulation to answer this question. Follow the four-step process.

28. **Scrabble** Refer to Exercise 20. About 3% of the time, the first player in Scrabble can "bingo" by playing all 7 tiles on the first turn. Should we be surprised if it takes 30 or more games for this to happen? Design and carry out a simulation to answer this question. Follow the four-step process.

29. **Random assignment** Researchers recruited 20 volunteers—8 men and 12 women—to take part in an experiment. They randomly assigned the subjects into two groups of 10 people each. To their surprise, 6 of the 8 men were randomly assigned to the same treatment. Should they be surprised? Design and carry out a simulation to estimate the probability that the random assignment puts 6 or more men in the same group. Follow the four-step process.

30. **Taking the train** According to New Jersey Transit, the 8:00 A.M. weekday train from Princeton to New York City has a 90% chance of arriving on time. To test this claim, an auditor chooses 6 weekdays at random during a month to ride this train. The train arrives late on 2 of those days. Does the auditor have convincing evidence that the company's claim isn't true? Design and carry out a simulation to estimate the probability that a train with a 90% chance of arriving on time each day would be late on 2 or more of 6 days. Follow the four-step process.

*Multiple choice: Select the best answer for Exercises 31 to 36.*

31. You read in a book about bridge that the probability that each of the four players is dealt exactly one ace is about 0.11. This means that

(a) in every 100 bridge deals, each player has one ace exactly 11 times.

(b) in 1 million bridge deals, the number of deals on which each player has one ace will be exactly 110,000.

(c) in a very large number of bridge deals, the percent of deals on which each player has one ace will be very close to 11%.

(d) in a very large number of bridge deals, the average number of aces in a hand will be very close to 0.11.

(e) If each player gets an ace in only 2 of the first 50 deals, then each player should get an ace in more than 11% of the next 50 deals.

32. If I toss a fair coin five times and the outcomes are TTTTT, then the probability that tails appears on the next toss is

(a) 0.5.   (c) greater than 0.5.   (e) 1.

(b) less than 0.5.   (d) 0.

Exercises 33 to 35 refer to the following setting. A basketball player claims to make 47% of her shots from the field. We want to simulate the player taking sets of 10 shots, assuming that her claim is true.

33. To simulate the number of makes in 10 shot attempts, you would perform the simulation as follows:

(a) Use 10 random one-digit numbers, where 0–4 are a make and 5–9 are a miss.

(b) Use 10 random two-digit numbers, where 00–46 are a make and 47–99 are a miss.

(c) Use 10 random two-digit numbers, where 00–47 are a make and 48–99 are a miss.

(d) Use 47 random one-digit numbers, where 0 is a make and 1–9 are a miss.

(e) Use 47 random two-digit numbers, where 00–46 are a make and 47–99 are a miss.

34. Twenty-five repetitions of the simulation were performed. The simulated number of makes in each set of 10 shots was recorded on the dotplot below. What is the approximate probability that a 47% shooter makes 5 or more shots in 10 attempts?

Number of Made Shots
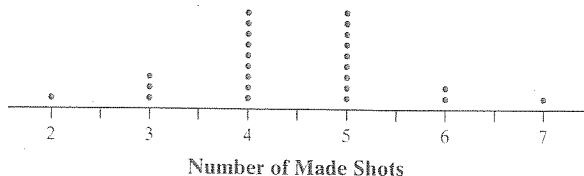
(a) 5/10 (b) 3/10 (c) 12/25 (d) 3/25 (e) 47/100

35. Suppose this player attempts 10 shots in a game and only makes 3 of them. Does this provide convincing evidence that she is less than a 47% shooter?

(a) Yes, because 3/10 (30%) is less than 47%.

(b) Yes, because she never made 47% of her shots in the simulation.

(d) No, because the simulation was only repeated 25 times.

(e) No, because the distribution is approximately symmetric.

36. Ten percent of U.S. households contain 5 or more people. You want to simulate choosing a household at random and recording "Yes" if it contains 5 or more people. Which of these are correct assignments of digits for this simulation?

(a) Odd = Yes; Even = No

(b) 0 = Yes; 1–9 = No

(c) 0–5 = Yes; 6–9 = No

(d) 0–4 = Yes; 5–9 = No

(e) None of these

37. **Are you feeling stressed?** (4.1) A Gallup Poll asked whether people experienced stress "a lot of the day yesterday." About 41 percent said they did. Gallup's report said, "Results are based on telephone interviews conducted … Jan. 1–Dec. 31, 2012, with a random sample of 353,564 adults aged 18 and older."[5]

(a) Identify the population and the sample.

(b) Explain how undercoverage could lead to bias in this survey.

38. **Waiting to park** (1.3, 4.2) Do drivers take longer to leave their parking spaces when someone is waiting? Researchers hung out in a parking lot and collected some data. The graphs and numerical summaries below display information about how long it took drivers to exit their spaces.

(a) Write a few sentences comparing these distributions.

(b) Can we conclude that having someone waiting causes drivers to leave their spaces more slowly? Why or why not?

Someone waiting?
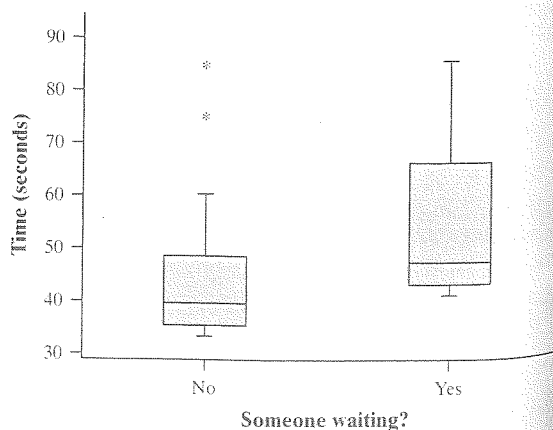
Descriptive Statistics: Time

Waiting n Mean StDev Minimum Q_1 Median Q_3 Maximum