CHAPTER 9

Discrete Mathematics

As the use of cellular telephones, modems, pagers, and fax machines has grown in recent years, the increasing demand for unique telephone numbers has necessitated the creation of new area codes in many areas of the United States. Counting the number of possible telephone numbers in a given area code is a *combinatorial* problem, and such problems are solved using the techniques of *discrete* mathematics. See page 709 for more on the subject of telephone area codes.

Chapter 9 Overview

The branches of mathematics known broadly as algebra, analysis, and geometry come together so beautifully in calculus that it has been difficult over the years to squeeze other mathematics into the curriculum. Consequently, many worthwhile topics like probability and statistics, combinatorics, graph theory, and numerical analysis that could easily be introduced in high school are (for many students) either first seen in college electives or else never seen at all. This situation is gradually changing as the applications of noncalculus mathematics become increasingly more important in the modern, computerized, data-driven workplace. Therefore, besides introducing important topics like sequences and series and the Binomial Theorem, this chapter will touch on some other discrete topics that might prove useful to you in the near future.

9.1 BASIC COMBINATORICS

What you'll learn about

- Discrete Versus Continuous
- The Importance of Counting
- The Multiplication Principle of Counting
- Permutations
- Combinations
- Subsets of an *n*-Set

. . . and why

Counting large sets is easy if you know the correct formula.

Discrete Versus Continuous

A point has no length and no width, and yet intervals on the real line—which are made up of these dimensionless points—have length! This little mystery illustrates the distinction between *continuous* and *discrete* mathematics. Any interval (a, b) contains a **continuum** of real numbers, which is why you can zoom in on an interval forever and there will still be an interval there. Calculus concepts like limits and continuity depend on the mathematics of the continuum. In *discrete* mathematics, we are concerned with properties of numbers and algebraic systems that do not depend on that kind of analysis. Many of them are related to the first kind of mathematics that most of us ever did, namely counting. Counting is what we will do for the rest of this section.

The Importance of Counting

We begin with a relatively simple counting problem.

EXAMPLE 1 Arranging three objects in order

In how many different ways can three distinguishable objects be arranged in order?

SOLUTION It is not difficult to list all the possibilities. If we call the objects A, B, and C, the different orderings are: *ABC, ACB, BAC, BCA, CAB,* and *CBA*. A good way to visualize the six choices is with a *tree diagram*, as in Figure 9.1. Notice that we have three choices for the first letter. Then, branching off each of those three choices are two choices for the second letter. Finally, branching off each of the $3 \times 2 = 6$ branches formed so far is one choice for the third letter. By beginning at the "root" of the tree, we can proceed to the right along any of the $3 \times 2 \times 1 = 6$ branches and get a different ordering each time. We conclude that there are six ways to arrange three distinguishable objects in order.

Now try Exercise 3.

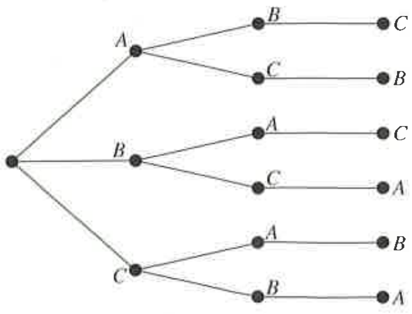


FIGURE 9.1 A tree diagram for ordering the letters *ABC*. (Example 1)

Scientific studies will usually manipulate one or more **explanatory** variables and observe the effect of that manipulation on one or more **response** variables. The key to understanding the significance of the effect is to know what is likely to occur *by chance alone*, and that often depends on counting. For example, Exploration 1 shows a real-world application of Example 1 above.

EXPLORATION 1 Questionable Product Claims

A salesman for a copying machine company is trying to convince a client to buy his $2000 machine instead of his competitor's $5000 machine. To make his point, he lines up an original document, a copy made by his machine, and a copy made by the more expensive machine on a table and asks 60 office workers to identify which is which. To everyone's surprise, not a single worker identifies all three correctly. The salesman states triumphantly that this proves that all three documents look the same to the naked eye and that therefore the client should buy his company's less expensive machine.

What do you think?

1. Each worker is essentially being asked to put the three documents in the correct order. How many ways can the three documents be ordered?
2. Suppose all three documents really *do* look alike. What fraction of the workers would you expect to put them into the correct order by chance alone?
3. If zero people out of 60 put the documents in the correct order, should we conclude that "all three documents look the same to the naked eye"?
4. Can you suggest a more likely conclusion that we might draw from the results of the salesman's experiment?

The Multiplication Principle of Counting

You would not want to draw the tree diagram for ordering five objects (*ABCDE*), but you should be able to see in your mind that it would have $5 \times 4 \times 3 \times 2 \times 1 = 120$ branches. A tree diagram is a geometric visualization of a fundamental counting principle known as the *Multiplication Principle*.

Multiplication Principle of Counting

If a procedure P has a sequence of stages $S_1, S_2, \ldots, S_n$ and if

$$S_1 \text{ can occur in } r_1 \text{ ways,}$$
$$S_2 \text{ can occur in } r_2 \text{ ways,}$$
$$\vdots$$
$$S_n \text{ can occur in } r_n \text{ ways,}$$

then the number of ways that the procedure P can occur is the product

$$r_1 r_2 \cdots r_n.$$

It is important to be mindful of how the choices at each stage are affected by the choices at preceding stages. For example, when choosing an order for the letters *ABC* we have 3 choices for the first letter, but only 2 choices for the second and 1 for the third.

EXAMPLE 2 Using the multiplication principle

The Tennessee license plates shown here consists of three letters of the alphabet followed by three numerical digits (0 through 9). Find the number of different license plates that could be formed

(a) if there is no restriction on the letters or digits that can be used;

(b) if no letter or digit can be repeated.

SOLUTION Consider each license plate as having six blanks to be filled in: three letters followed by three numerical digits.

(a) If there are no restrictions on letters or digits, then we can fill in the first blank 26 ways, the second blank 26 ways, the third blank 26 ways, the fourth blank 10 ways, the fifth blank 10 ways, and the sixth blank 10 ways. By the Multiplication Principle, we can fill in all six blanks in $26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17{,}576{,}000$ ways. There are 17,576,000 possible license plates with no restrictions on letters or digits.

(b) If no letter or digit can be repeated, then we can fill in the first blank 26 ways, the second blank 25 ways, the third blank 24 ways, the fourth blank 10 ways, the fifth blank 9 ways, and the sixth blank 8 ways. By the Multiplication Principle, we can fill in all six blanks in $26 \times 25 \times 24 \times 10 \times 9 \times 8 = 11{,}232{,}000$ ways. There are 11,232,000 possible license plates with no letters or digits repeated.

Now try Exercise 5.

LICENSE PLATE RESTRICTIONS

Although prohibiting repeated letters and digits as in Example 2 would make no practical sense (why rule out more than 6 million possible plates for no good reason?), states do impose some restrictions on license plates. They rule out certain letter progressions that could be considered obscene or offensive.

Permutations

One important application of the Multiplication Principle of Counting is to count the number of ways that a set of n objects (called an ***n*-set**) can be arranged in order. Each such ordering is called a **permutation** of the set. Example 1 showed that there are $3! = 6$ permutations of a 3-set. In fact, if you understood the tree diagram, you can probably guess how many permutations there are of an n-set.

FACTORIALS

If n is a positive integer, the symbol $n!$ (read "n factorial") represents the product $n(n-1)(n-2)(n-3)\cdots 2 \cdot 1$. We also define $0! = 1$.

Permutations of an n-set

There are $n!$ permutations of an n-set.

Usually the elements of a set are distinguishable from one another, but we can adjust our counting when they are not, as we see in Example 3.

EXAMPLE 3 Distinguishable permutations

Count the number of different 9-letter "words" (don't worry about whether they're in the dictionary) that can be formed using the letters in each word.

(a) DRAGONFLY **(b)** BUTTERFLY **(c)** BUMBLEBEE

SOLUTION

(a) Each permutation of the 9 letters forms a different word. There are $9! = 362{,}880$ such permutations.

(b) There are also 9! permutations of these letters, but a simple permutation of the two T's does not result in a new word. We correct for the overcount by dividing by 2!. There are $\frac{9!}{2!} = 181{,}440$ *distinguishable* permutations of the letters in BUTTERFLY.

(c) Again there are 9! permutations, but the three B's are indistinguishable, as are the three E's, so we divide by 3! twice to correct for the overcount. There are $\frac{9!}{3!3!} = 10{,}080$ distinguishable permutations of the letters in BUMBLEBEE.

Now try Exercise 9.

Distinguishable Permutations

There are $n!$ distinguishable permutations of an n-set containing n distinguishable objects.

If an n-set contains n_1 objects of a first kind, n_2 objects of a second kind, and so on, with $n_1 + n_2 + \cdots + n_k = n$, then the number of distinguishable permutations of the n-set is

$$\frac{n!}{n_1!n_2!n_3!\cdots n_k!}.$$

In many counting problems, we are interested in using n objects to fill r blanks in order, where $r < n$. These are called **permutations of n objects taken r at a time**. The procedure for counting them is the same; only this time we run out of blanks before we run out of objects.

The first blank can be filled in n ways, the second in $n - 1$ ways, and so on until we come to the rth blank, which can be filled in $n - (r - 1)$ ways.

By the Multiplication Principle, we can fill in all r blanks in $n(n-1)(n-2)\cdots(n-r+1)$ ways. This expression can be written in a more compact (but less easily computed) way as $n!/(n-r)!$.

PERMUTATIONS ON A CALCULATOR

Most modern calculators have an ${}_nP_r$ selection built in. They also compute factorials, but remember that factorials get very large. If you want to count the number of permutations of 90 objects taken 5 at a time, be sure to use the ${}_nP_r$ function. The expression 90!/85! is likely to lead to an overflow error.

Permutation Counting Formula

The number of permutations of n objects taken r at a time is denoted ${}_nP_r$ and is given by

$${}_nP_r = \frac{n!}{(n-r)!} \quad \text{for} \quad 0 \le r \le n.$$

If $r > n$, then ${}_nP_r = 0$.

Notice that ${}_nP_n = n!/(n-n)! = n!/0! = n!/1 = n!$, which we have already seen is the number of permutations of a complete set of n objects. This is why we define $0! = 1$.

EXAMPLE 4 Counting permutations

Evaluate each expression without a calculator.

(a) ${}_6P_4$ **(b)** ${}_{11}P_3$ **(c)** ${}_nP_3$

SOLUTION

(a) By the formula, ${}_6P_4 = 6!/(6-4)! = 6!/2! = (6 \cdot 5 \cdot 4 \cdot 3 \cdot 2!)/2! = 6 \cdot 5 \cdot 4 \cdot 3 = 360$.

(b) Although you could use the formula again, you might prefer to apply the Multiplication Principle directly. We have 11 objects and 3 blanks to fill:

$${}_{11}P_3 = 11 \cdot 10 \cdot 9 = 990.$$

(c) This time it is definitely easier to use the Multiplication Principle. We have n objects and 3 blanks to fill; so assuming $n \ge 3$,

$${}_nP_3 = n(n-1)(n-2).$$

Now try Exercise 15.

EXAMPLE 5 Applying permutations

Sixteen actors answer a casting call to try out for roles as dwarfs in a production of *Snow White and the Seven Dwarfs*. In how many different ways can the director cast the seven roles?

SOLUTION The 7 different roles can be thought of as 7 blanks to be filled, and we have 16 actors with which to fill them. The director can cast the roles in ${}_{16}P_7 = 57{,}657{,}600$ ways.

Now try Exercise 12.

Combinations

When we count permutations of n objects taken r at a time, we consider different orderings of the same r selected objects as being different permutations.

In many applications we are only interested in the ways to *select* the r objects, regardless of the order in which we arrange them. These unordered selections are called **combinations of n objects taken r at a time**.

Combination Counting Formula

The number of combinations of n objects taken r at a time is denoted ${}_nC_r$ and is given by

$$ {}_nC_r = \frac{n!}{r!(n-r)!} \quad \text{for} \quad 0 \le r \le n. $$

If $r > n$, then ${}_nC_r = 0$.

We can verify the ${}_nC_r$ formula with the Multiplication Principle. Since every permutation can be thought of as an *unordered* selection of r objects *followed* by a particular *ordering* of the objects selected, the Multiplication Principle gives ${}_nP_r = {}_nC_r \cdot r!$.

Therefore

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{1}{r!} \cdot \frac{n!}{(n-r)!} = \frac{n!}{r!(n-r)!}. $$

A WORD ON NOTATION

Some textbooks use $P(n, r)$ instead of ${}_nP_r$ and $C(n, r)$ instead of ${}_nC_r$. Much more common is the notation $\binom{n}{r}$ for ${}_nC_r$. Both $\binom{n}{r}$ and ${}_nC_r$ are often read "n choose r."

COMBINATIONS ON A CALCULATOR

Most modern calculators have an *nCr* selection built in. As with permutations, it is better to use the *nCr* function than to use the formula $\frac{n!}{r!(n-r)!}$, as the individual factorials can get too large for the calculator.

EXAMPLE 6 Distinguishing combinations from permutations

In each of the following scenarios, tell whether permutations (ordered) or combinations (unordered) are being described.

(a) A president, vice-president, and secretary are chosen from a 25-member garden club.

(b) A cook chooses 5 potatoes from a bag of 12 potatoes to make a potato salad.

(c) A teacher makes a seating chart for 22 students in a classroom with 30 desks.

SOLUTION

(a) Permutations. Order matters because it matters who gets which office.

(b) Combinations. The salad is the same no matter what order the potatoes are chosen.

(c) Permutations. A different ordering of students in the same seats results in a different seating chart.

Notice that once you know what is being counted, getting the correct number is easy with a calculator. The number of possible choices in the scenarios above are: (a) ${}_{25}P_3 = 13{,}800$, (b) ${}_{12}C_5 = 792$, and (c) ${}_{30}P_{22} \approx 6.5787 \times 10^{27}$.

Now try Exercise 19.

EXAMPLE 7 Counting combinations

In the Miss America pageant, 51 contestants must be narrowed down to 10 finalists who will compete on national television. In how many possible ways can the ten finalists be selected?

SOLUTION Notice that the *order* of the finalists does not matter at this phase; all that matters is which women are selected. So we count combinations rather than permutations.

$$_{51}C_{10} = \frac{51!}{10!41!} = 12{,}777{,}711{,}870.$$

The 10 finalists can be chosen in 12,777,711,870 ways.

Now try Exercise 27.

EXAMPLE 8 Picking lottery numbers

The Georgia Lotto requires winners to pick 6 integers between 1 and 46. The order in which you select them does not matter; indeed, the lottery tickets are always printed with the numbers in ascending order. How many different lottery tickets are possible?

SOLUTION There are $_{46}C_6 = 9{,}366{,}819$ possible lottery tickets of this type. (That's more than enough different tickets for every person in the state of Georgia!)

Now try Exercise 29.

Subsets of an *n*-Set

As a final application of the counting principle, consider the pizza topping problem.

EXAMPLE 9 Selecting pizza toppings

Armando's Pizzeria offers patrons any combination of up to 10 different toppings: pepperoni, mushroom, sausage, onion, green pepper, bacon, prosciutto, black olive, green olive, and anchovies. How many different pizzas can be ordered

(a) if we can choose any three toppings?

(b) if we can choose any number of toppings (0 through 10)?

SOLUTION

(a) Order does not matter (for example, the sausage-pepperoni-mushroom pizza is the same as the pepperoni-mushroom-sausage pizza), so the number of possible pizzas is $_{10}C_3 = 120$.

(b) We could add up all the numbers of the form $_{10}C_r$ for $r = 0, 1, \ldots, 10$, but there is an easier way to count the possibilities. Consider the ten options to be lined up as in the statement of the problem. In considering each option, we have two choices: yes or no. (For example, the pepperoni-mushroom-sausage pizza would correspond to the sequence YYYNNNNNNN.) By the

Multiplication Principle, the number of such sequences is $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 1024$, which is the number of possible pizzas.

Now try Exercise 37.

The solution to Example 9b suggests a general rule that will be our last counting formula of the section.

Formula for Counting Subsets of an *n*-Set

There are 2^n subsets of a set with n objects (including the empty set and the entire set).

EXAMPLE 10 Analyzing an advertised claim

A national hamburger chain used to advertise that it fixed its hamburgers "256 ways," since patrons could order whatever toppings they wanted. How many toppings must have been available?

SOLUTION We need to solve the equation $2^n = 256$ for n. We could solve this easily enough by trial and error, but we will solve it with logarithms just to keep the method fresh in our minds.

$$\begin{aligned} 2^n &= 256 \\ \log 2^n &= \log 256 \\ n \log 2 &= \log 256 \\ n &= \frac{\log 256}{\log 2} \\ n &= 8 \end{aligned}$$

There must have been 8 toppings from which to choose.

Now try Exercise 39.

Why are there not 1000 possible Area Codes?

While there are 1000 three-digit numbers between 000 and 999, not all of them are available for use as area codes. For example, area codes cannot begin with 0 or 1, and numbers of the form *abb* have been reserved for other purposes.

PROBLEM: There are 680 three-digit numbers that are available for use as area codes in North America. As of April 2002, 305 of them were actually being used (*Source: www.nanpa.com*). How many additional three-digit area codes are available for use? Within a given area code, how many unique telephone numbers are theoretically possible?

SOLUTION: There are $680 - 305 = 375$ additional area codes available. Within a given area code, each telephone number has seven digits chosen from the ten digits 0 through 9. Since each digit can theoretically be any of 10 numbers, there are

$$10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 10^7 = 10{,}000{,}000$$

different telephone numbers possible within a given area code.

Putting these two results together, we see that the unused area codes in April 2002 represented an additional 3.75 billion possible telephone numbers!

QUICK REVIEW 9.1

In Exercises 1–10, give the number of objects described. In some cases you might have to do a little research or ask a friend.

1. The number of cards in a standard deck
2. The number of cards of each suit in a standard deck
3. The number of faces on a cubical die
4. The number of possible totals when two dice are rolled
5. The number of vertices of a decagon
6. The number of musicians in a string quartet
7. The number of players on a soccer team
8. The number of prime numbers between 1 and 10, inclusive
9. The number of squares on a chessboard
10. The number of cards in a contract bridge hand

SECTION 9.1 EXERCISES

In Exercises 1–4, count the number of ways that each procedure can be done.

1. Line up three people for a photograph.
2. Prioritize four pending jobs from most to least important.
3. Arrange five books from left to right on a bookshelf.
4. Award ribbons for 1st place to 5th place to the top five dogs in a dog show.
5. **Homecoming King and Queen** There are four candidates for homecoming queen and three candidates for king. How many king-queen pairs are possible?
6. **Possible Routes** There are three roads from town A to town B and four roads from town B to town C. How many different routes are there from A to C by way of B?
7. **Permuting Letters** How many 9-letter "words" (not necessarily in any dictionary) can be formed from the letters of the word LOGARITHM? (Curiously, one such arrangement spells another word related to mathematics. Can you name it?)
8. **Three-Letter Crossword Entries** Excluding J, Q, X, and Z, how many 3-letter crossword puzzle entries can be formed that contain no repeated letters? (It has been conjectured that all of them have appeared in puzzles over the years, sometimes with painfully contrived definitions.)
9. **Permuting Letters** How many distinguishable 11-letter "words" can be formed using the letters in MISSISSIPPI?
10. **Permuting Letters** How many distinguishable 11-letter "words" can be formed using the letters in CHATTANOOGA?
11. **Electing Officers** The 13 members of the East Brainerd Garden Club are electing a President, Vice-President, and Secretary from among their members. How many different ways can this be done?
12. **City Government** From among 12 projects under consideration, the mayor must put together a prioritized (that is, ordered) list of 6 projects to submit to the city council for funding. How many such lists can be formed?

In Exercises 13–18, evaluate each expression without a calculator. Then check with your calculator to see if your answer is correct.

13. $4!$
14. $(3!)(0!)$
15. ${}_6P_2$
16. ${}_9P_2$
17. ${}_{10}C_7$
18. ${}_{10}C_3$

In Exercises 19–22, tell whether permutations (ordered) or combinations (unordered) are being described.

19. 13 cards are selected from a deck of 52 to form a bridge hand.
20. 7 digits are selected (without repetition) to form a telephone number.
21. 4 students are selected from the senior class to form a committee to advise the cafeteria director about food.
22. 4 actors are chosen to play the Beatles in a film biography.
23. **License Plates** How many different license plates begin with two digits, followed by two letters and then three digits if no letters or digits are repeated?
24. **License Plates** How many different license plates consist of five symbols, either digits or letters?
25. **Tumbling Dice** Suppose that two dice, one red and one green, are rolled. How many different outcomes are possible for the pair of dice?
26. **Coin Toss** How many different sequences of heads and tails are there if a coin is tossed 10 times?
27. **Forming Committees** A 3-woman committee is to be elected from a 25-member sorority. How many different committees can be elected?

28. **Straight Poker** In the original version of poker known as "straight" poker, a five-card hand is dealt from a standard deck of 52. How many different straight poker hands are possible?

29. **Buying Discs** Juan has money to buy only three of the 48 compact discs available. How many different sets of discs can he purchase?

30. **Coin Toss** A coin is tossed 20 times and the heads and tails sequence is recorded. From among all the possible sequences of heads and tails, how many have exactly seven heads?

31. **Drawing Cards** How many different 13-card hands include the ace and king of spades?

32. **Job Interviews** The head of the personnel department interviews eight people for three identical openings. How many different groups of three can be employed?

33. **Scholarship Nominations** Six seniors at Rydell High School meet the qualifications for a competitive honor scholarship at a major university. The university allows the school to nominate up to three candidates, and the school always nominates at least one. How many different choices could the nominating committee make?

34. **Pu-pu Platters** A Chinese restaurant will make a Pu-pu platter "to order" containing any one, two, or three selections from its appetizer menu. If the menu offers five different appetizers, how many different platters could be made?

35. **Yahtzee** In the game of Yahtzee, five dice are tossed simultaneously. How many outcomes can be distinguished if all the dice are different colors?

36. **Indiana Jones and the Final Exam** Professor Indiana Jones gives his class 20 study questions, from which he will select 8 to be answered on the final exam. How many ways can he select the questions?

37. **Salad Bar** Mary's lunch always consists of a full plate of salad from Ernestine's salad bar. She always takes equal amounts of each salad she chooses, but she likes to vary her selections. If she can choose from among 9 different salads, how many essentially different lunches can she create?

38. **Buying a New Car** A new car customer has to choose from among 3 models, each of which comes in 4 exterior colors, 3 interior colors, and with any combination of up to 6 optional accessories. How many essentially different ways can the customer order the car?

39. **Pizza Possibilities** Luigi sells one size of pizza, but he claims that his selection of toppings allows for "more than 4000 different choices." What is the smallest number of toppings Luigi could offer?

40. **Proper Subsets** A subset of set A is called *proper* if it is neither the empty set nor the entire set A. How many proper subsets does an n-set have?

41. **True-False Tests** How many different answer keys are possible for a 10-question True-False test?

42. **Multiple-Choice Tests** How many different answer keys are possible for a 10-question multiple-choice test in which each question leads to choice a, b, c, d, or e?

Standardized Test Questions

43. **True or False** If a and b are positive integers such that $a + b = n$, then $\binom{n}{a} = \binom{n}{b}$. Justify your answer.

44. **True or False** If a, b, and n are integers such that $a < b < n$, then $\binom{n}{a} < \binom{n}{b}$. Justify your answer.

You may use a graphing calculator when evaluating Exercises 45–48.

45. Lunch at the Gritsy Palace consists of an entrée, two vegetables, and a dessert. If there are four entrées, six vegetables, and six desserts from which to choose, how many essentially different lunches are possible?
(a) 16
(b) 25
(c) 144
(d) 360
(e) 720

46. How many different ways can the judges choose 5th to 1st places from ten Miss America finalists?
(a) 50
(b) 120
(c) 252
(d) 30,240
(e) 3,628,800

47. Assuming r and n are positive integers with $r < n$, which of the following numbers does *not* equal 1?
(a) $(n - n)!$
(b) ${}_nP_n$
(c) ${}_nC_n$
(d) $\binom{n}{n}$
(e) $\binom{n}{r} \div \binom{n}{n-r}$

48. An organization is electing 3 new board members by approval voting. Members are given ballots with the names of 5 candidates and are allowed to check off the names of all candidates whom they would approve (which could be none, or even all five). The three candidates with the most checks overall are elected. In how many different ways can a member fill out the ballot?

(a) 10

(b) 20

(c) 32

(d) 125

(e) 243

Explorations

49. Group Activity For each of the following numbers, make up a counting problem that has the number as its answer.

(a) ${}_{52}C_3$

(b) ${}_{12}C_3$

(c) ${}_{25}P_{11}$

(d) 2^5

(e) $3 \cdot 2^{10}$

50. Writing to Learn You have a fresh carton containing one dozen eggs and you need to choose two for breakfast. Give a counting argument based on this scenario to explain why ${}_{12}C_2 = {}_{12}C_{10}$.

51. Factorial Riddle The number 50! ends in a string of consecutive 0's.

(a) How many 0's are in the string?

(b) How do you know?

52. Group Activity Diagonals of a Regular Polygon In Exploration 1 of Section 1.6, you reasoned from data points and quadratic regression that the number of diagonals of a regular polygon with n vertices was $(n^2 - 3n)/2$.

(a) Explain why the number of segments connecting all pairs of vertices is ${}_nC_2$.

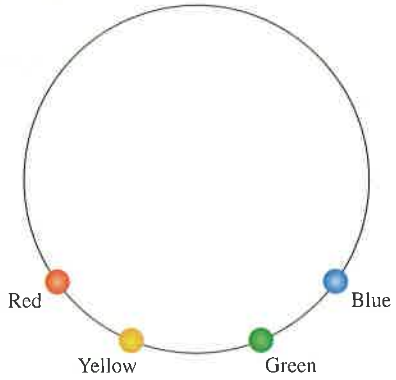
(b) Use the result from (a) to prove that the number of diagonals is $(n^2 - 3n)/2$.

Extending the Ideas

53. Writing to Learn Suppose that a chain letter (illegal if money is involved) is sent to five people the first week of the year. Each of these five people sends a copy of the letter to five more people during the second week of the year. Assume that everyone who receives a letter participates. Explain how you know with certainty that someone will receive a second copy of this letter later in the year.

54. A Round Table How many different seating arrangements are possible for 4 people sitting around a round table?

55. Colored Beads Four beads—red, blue, yellow, and green—are arranged on a string to make a simple necklace as shown in the figure. How many arrangements are possible?



56. Casting a Play A director is casting a play with two female leads and wants to have a chance to audition the actresses two at a time to get a feeling for how well they would work together. His casting director and his administrative assistant both prepare charts to show the amount of time that would be required, depending on the number of actresses who come to the audition. Which time chart is more reasonable, and why?

Number who audition	Time required (minutes)
3	10
6	45
9	110
12	200
15	320

Number who audition	Time required (minutes)
3	10
6	30
9	60
12	100
15	150

57. Bridge Around the World Suppose that a contract bridge hand is dealt somewhere in the world every second. What is the fewest number of years required for every possible bridge hand to be dealt? (See Quick Review Exercise 10.)

58. Basketball Lineups Each NBA basketball team has 12 players on its roster. If each coach chooses 5 starters without regard to position, how many different sets of 10 players can start when two given teams play a game?

9.2 THE BINOMIAL THEOREM

What you'll learn about

- Powers of Binomials
- Pascal's Triangle
- The Binomial Theorem
- Factorial Identities

. . . and why

The Binomial Theorem is a marvelous study in combinatorial patterns.

Powers of Binomials

Many important mathematical discoveries have begun with the study of patterns. In this chapter, we want to introduce an important polynomial theorem called the Binomial Theorem, for which we will set the stage by observing some patterns.

If you expand $(a + b)^n$ for $n = 0, 1, 2, 3, 4,$ and 5, here is what you get:

$$\begin{aligned}
(a+b)^0 &= 1\\
(a+b)^1 &= 1a^1b^0 + 1a^0b^1\\
(a+b)^2 &= 1a^2b^0 + 2a^1b^1 + 1a^0b^2\\
(a+b)^3 &= 1a^3b^0 + 3a^2b^1 + 3a^1b^2 + 1a^0b^3\\
(a+b)^4 &= 1a^4b^0 + 4a^3b^1 + 6a^2b^2 + 4a^1b^3 + 1a^0b^4\\
(a+b)^5 &= 1a^5b^0 + 5a^4b^1 + 10a^3b^2 + 10a^2b^3 + 5a^1b^4 + 1a^0b^5
\end{aligned}$$

Can you observe the patterns and predict what the expansion of $(a + b)^6$ will look like? You can probably predict the following:

1. The powers of a will decrease from 6 to 0 by 1's.
2. The powers of b will increase from 0 to 6 by 1's.
3. The first two coefficients will be 1 and 6.
4. The last two coefficients will be 6 and 1.

At first you might not see the pattern that would enable you to find the other so-called *binomial coefficients*, but you should see it after the following Exploration.

EXPLORATION 1 Exploring the Binomial Coefficients

1. Compute ${}_3C_0$, ${}_3C_1$, ${}_3C_2$, and ${}_3C_3$. Where can you find these numbers in the binomial expansions above?
2. The expression $4\ {}_nC_r\ \{0, 1, 2, 3, 4\}$ tells the calculator to compute ${}_4C_r$ for each of the numbers $r = 0, 1, 2, 3, 4$ and display them as a list. Where can you find these numbers in the binomial expansions above?
3. Compute $5\ {}_nC_r\ \{0, 1, 2, 3, 4, 5\}$. Where can you find these numbers in the binomial expansions above?

By now you are probably ready to conclude that the binomial coefficients in the expansion of $(a + b)^n$ are just the values of ${}_nC_r$ for $r = 0, 1, 2, 3, 4, \ldots, n$. We also hope you are wondering *why* this is true.

The expansion of

$$(a+b)^n = \underbrace{(a+b)(a+b)(a+b)\cdots(a+b)}_{n \text{ factors}}$$

consists of all possible products that can be formed by taking one letter (either a or b) from each factor $(a + b)$. The number of ways to form the product $a^r b^{n-r}$ is exactly the same as the number of ways to choose r factors to contribute an a, since the rest of the factors will obviously contribute a b. The number of ways to choose r factors from n factors is ${}_nC_r$.

TABLE TRICK

You can also use the table display to show binomial coefficients. For example, let Y1 = 5 ${}_nC_r X$, and set TblStart = 0 and $\Delta Tbl = 1$ to display the binomial coefficients for $(a + b)^5$.

Definition Binomial Coefficient

The binomial coefficients that appear in the expansion of $(a + b)^n$ are the values of ${}_nC_r$ for $r = 0, 1, 2, 3, \ldots, n$.

A classical notation for ${}_nC_r$, especially in the context of binomial coefficients, is $\binom{n}{r}$. Both notations are read "n choose r."

EXAMPLE 1 Using ${}_nC_r$ to expand a binomial

Expand $(a + b)^5$, using a calculator to compute the binomial coefficients.

SOLUTION Enter 5 ${}_nC_r$ {0, 1, 2, 3, 4, 5} into the calculator to find the binomial coefficients for $n = 5$. The calculator returns the list {1, 5, 10, 10, 5, 1}. Using these coefficients, we construct the expansion:

$$(a+b)^5 = 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5.$$

Now try Exercise 3.

Pascal's Triangle

If we eliminate the plus signs and the powers of the variables a and b in the "triangular" array of binomial coefficients with which we began this section, we get:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
\iddots & & & & & \vdots & & & & & \ddots
\end{array}$$

THE NAME GAME

The fact that Pascal's triangle was not discovered by Pascal is ironic, but hardly unusual in the annals of mathematics. We mentioned in Chapter 5 that Heron did not discover Heron's formula, and Pythagoras did not even discover the Pythagorean theorem. The history of calculus is filled with similar injustices.

This is called **Pascal's triangle** in honor of Blaise Pascal (1623–1662), who used it in his work but certainly did not discover it. It appeared in 1303 in a Chinese text, the *Precious Mirror*, by Chu Shih-chieh, who referred to it even then as a "diagram of the old method for finding eighth and lower powers."

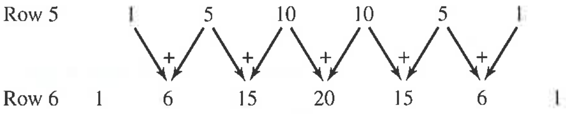
For convenience, we refer to the top "1" in Pascal's triangle as row 0. That allows us to associate the numbers along row n with the expansion of $(a + b)^n$.

Pascal's triangle is so rich in patterns that people still write about them today. One of the simplest patterns is the one that we use for getting from one row to the next, as in the following example.

EXAMPLE 2 Extending Pascal's triangle

Show how row 5 of Pascal's triangle can be used to obtain row 6, and use the information to write the expansion of $(x + y)^6$.

SOLUTION The two outer numbers of every row are 1's. Each number between them is the sum of the two numbers immediately above it. So row 6 can be found from row 5 as follows:



These are the binomial coefficients for $(x + y)^6$, so

$$(x + y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6.$$

Now try Exercise 7.

The technique used in Example 1 to extend Pascal's triangle depends on the following recursion formula.

Recursion Formula for Pascal's Triangle

$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$ or, equivalently, ${}_nC_r = {}_{n-1}C_{r-1} + {}_{n-1}C_r$

Here's a counting argument to explain why it works. Suppose we are choosing r objects from n objects. As we have seen, this can be done in ${}_nC_r$ ways. Now identify one of the n objects with a special tag. How many ways can we choose r objects if the tagged object is among them? Well, we have $r - 1$ objects yet to be chosen from among the $n - 1$ that are untagged, so ${}_{n-1}C_{r-1}$. How many ways can we choose r objects if the tagged object is *not* among them? This time we must choose all r objects from among the $n - 1$ without tags, so ${}_{n-1}C_r$. Since our selection of r objects must either contain the tagged object or not contain it, ${}_{n-1}C_{r-1} + {}_{n-1}C_r$ counts all the possibilities with no overlap. Therefore, ${}_nC_r = {}_{n-1}C_{r-1} + {}_{n-1}C_r$.

It is not necessary to construct Pascal's triangle to find specific binomial coefficients, since we already have a formula for computing them: ${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$. This formula can be used to give an algebraic formula for the recursion formula above, but we will leave that as an exercise for the end of the section.

EXAMPLE 3 Computing binomial coefficients

Find the coefficient of x^{10} in the expansion of $(x + 2)^{15}$.

SOLUTION The only term in the expansion that we need to deal with is ${}_{15}C_{10}x^{10}2^5$. This is

$$\frac{15!}{10!5!} \cdot 2^5 \cdot x^{10} = 3003 \cdot 32 \cdot x^{10} = 96{,}096\, x^{10}.$$

The coefficient of x^{10} is 96,096. Now try Exercise 15.

THE BINOMIAL THEOREM IN Σ NOTATION

For those who are already familiar with summation notation, here is how the Binomial Theorem looks:

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r}b^r.$$

Those who are not familiar with this notation will learn about it in Section 9.4.

The Binomial Theorem

We now state formally the theorem about expanding powers of binomials, known as the Binomial Theorem. For tradition's sake, we will use the symbol $\binom{n}{r}$ instead of ${}_nC_r$.

The Binomial Theorem

For any positive integer n,

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + \binom{n}{n}b^n,$$

where

$$\binom{n}{r} = {}_nC_r = \frac{n!}{r!(n-r)!}.$$

EXAMPLE 4 Expanding a binomial

Expand $(2x - y^2)^4$.

SOLUTION We use the Binomial Theorem to expand $(a + b)^4$ where $a = 2x$ and $b = -y^2$.

$$\begin{aligned}(a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\\ (2x-y^2)^4 &= (2x)^4 + 4(2x)^3(-y^2) + 6(2x)^2(-y^2)^2\\ &\quad + 4(2x)(-y^2)^3 + (-y^2)^4\\ &= 16x^4 - 32x^3y^2 + 24x^2y^4 - 8xy^6 + y^8\end{aligned}$$

Now try Exercise 17.

Factorial Identities

Expressions involving factorials combine to give some interesting identities, most of them relying on the basic identities shown below (actually two versions of the same identity).

Basic Factorial Identities

For any integer $n \geq 1$, $n! = n(n-1)!$

For any integer $n \geq 0$, $(n+1)! = (n+1)n!$

EXAMPLE 5 Proving an identity with factorials

Prove that $\binom{n+1}{2} - \binom{n}{2} = n$ for all integers $n \geq 2$.

SOLUTION

$$\binom{n+1}{2} - \binom{n}{2} = \frac{(n+1)!}{2!(n+1-2)!} - \frac{n!}{2!(n-2)!}$$ Combination counting formula

$$= \frac{(n+1)(n)(n-1)!}{2(n-1)!} - \frac{n(n-1)(n-2)!}{2(n-2)!}$$ Basic factorial identities

$$= \frac{n^2+n}{2} - \frac{n^2-n}{2}$$

$$= \frac{2n}{2}$$

$$= n$$

Now try Exercise 33.

QUICK REVIEW 9.2

(Prerequisite skill Section A.2)

In Exercises 1–10, use the distributive property to expand the binomial.

1. $(x + y)^2$

2. $(a + b)^2$

3. $(5x - y)^2$

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

5. $(3s + 2t)^2$

6. $(3p - 4q)^2$

7. $(u + v)^3$

8. $(b - c)^3$

9. $(2x - 3y)^3$

10. $(4m + 3n)^3$

SECTION 9.2 EXERCISES

In Exercises 1–4, expand the binomial using a calculator to find the binomial coefficients.

1. $(a + b)^4$

2. $(a + b)^6$

3. $(x + y)^7$

4. $(x + y)^{10}$

In Exercises 5–8, expand the binomial using Pascal's triangle to find the coefficients.

5. $(x + y)^3$

6. $(x + y)^5$

7. $(p + q)^8$

8. $(p + q)^9$

In Exercises 9–12, evaluate the expression by hand (using the formula) before checking your answer on a grapher.

9. $\binom{9}{2}$

10. $\binom{15}{11}$

11. $\binom{166}{166}$

12. $\binom{166}{0}$

In Exercises 13–16, find the coefficient of the given term in the binomial expansion.

13. $x^{11}y^3$ term, $(x + y)^{14}$

14. x^5y^8 term, $(x + y)^{13}$

15. x^4 term, $(x - 2)^{12}$

16. x^7 term, $(x - 3)^{11}$

In Exercises 17–20, use the Binomial Theorem to find a polynomial expansion for the function.

17. $f(x) = (x - 2)^5$

18. $g(x) = (x + 3)^6$

19. $h(x) = (2x - 1)^7$

20. $f(x) = (3x + 4)^5$

In Exercises 21–26, use the Binomial Theorem to expand each expression.

21. $(2x + y)^4$

22. $(2y - 3x)^5$

23. $(\sqrt{x} - \sqrt{y})^6$

24. $(\sqrt{x} + \sqrt{3})^4$

25. $(x^{-2} + 3)^5$

26. $(a - b^{-3})^7$

27. Determine the largest integer n for which your calculator will compute $n!$.

28. Determine the largest integer n for which your calculator will compute $\binom{n}{100}$.

29. Prove that $\binom{n}{1} = \binom{n}{n-1} = n$ for all integers $n \geq 1$.

30. Prove that $\binom{n}{r} = \binom{n}{n-r}$ for all integers $n \geq r \geq 0$.

31. Use the formula $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ to prove that $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$. (This is the pattern in Pascal's triangle that appears in Example 2.)

32. Find a counterexample to show that each statement is *false.*

(a) $(n + m)! = n! + m!$

(b) $(nm)! = n!m!$

33. Prove that $\binom{n}{2} + \binom{n+1}{2} = n^2$ for all integers $n \geq 2$.

34. Prove that $\binom{n}{n-2} + \binom{n+1}{n-1} = n^2$ for all integers $n \geq 2$.

Standardized Test Questions

35. **True or False** The coefficients in the polynomial expansion of $(x - y)^{50}$ alternate in sign. Justify your answer.

36. **True or False** The sum of any row of Pascal's triangle is an even integer. Justify your answer.

You may use a graphing calculator when evaluating Exercises 37–40.

37. What is the coefficient of x^4 in the expansion of $(2x + 1)^8$?

(a) 16

(b) 256

(c) 1120

(d) 1680

(e) 26,680

38. Which of the following numbers does *not* appear on row 10 of Pascal's Triangle?

(a) 1

(b) 5

(c) 10

(d) 120

(e) 252

39. The *sum* of the coefficients of $(3x - 2y)^{10}$ is

(a) 1

(b) 1024

(c) 58,025

(d) 59,049

(e) 9,765,625

40. $(x + y)^3 + (x - y)^3 =$

(a) 0

(b) $2x^3$

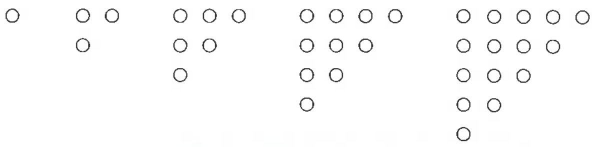
(c) $2x^3 - 2y^3$

(d) $2x^3 + 6xy^2$

(e) $6x^2y + 2y^3$

Explorations

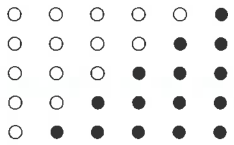
41. **Triangular Numbers** Numbers of the form $1 + 2 + \cdots + n$ are called **triangular numbers** because they count numbers in triangular arrays, as shown below:



(a) Compute the first 10 triangular numbers.

(b) Where do the triangular numbers appear in Pascal's triangle?

(c) **Writing to Learn** Explain why the diagram below shows that the *n*th triangular number can be written as $n(n + 1)/2$.



(d) Write the formula in (c) as a binomial coefficient. (This is why the triangular numbers appear as they do in Pascal's triangle.)

42. **Group Activity** **Exploring Pascal's Triangle** Break into groups of two or three. Just by looking at patterns in Pascal's triangle, guess the answers to the following questions. (It is easier to make a conjecture from a pattern than it is to construct a proof!)

(a) What positive integer appears the least number of times?

(b) What number appears the greatest number of times?

(c) Is there any positive integer that does *not* appear in Pascal's triangle?

(d) If you go along any row alternately adding and subtracting the numbers, what is the result?

(e) If p is a prime number, what do all the interior numbers along the pth row have in common?

(f) Which rows have all even interior numbers?

(g) Which rows have all odd numbers?

(h) What other patterns can you find? Share your discoveries with the other groups.

Extending the Ideas

43. Use the Binomial Theorem to prove that the sum of the entries along the nth row of Pascal's triangle is 2^n. That is,

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n.$$

[*Hint*: Use the Binomial Theorem to expand $(1 + 1)^n$.]

44. Use the Binomial Theorem to prove that the alternating sum along any row of Pascal's triangle is zero. That is,

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n\binom{n}{n} = 0.$$

45. Use the Binomial Theorem to prove that

$$\binom{n}{0} + 2\binom{n}{1} + 4\binom{n}{2} + \cdots + 2^n\binom{n}{n} = 3^n.$$

9.3 PROBABILITY

What you'll learn about

- Sample Spaces and Probability Functions
- Determining Probabilities
- Venn Diagrams and Tree Diagrams
- Conditional Probability
- Binomial Distributions

. . . and why

Everyone should know how mathematical the "laws of chance" really are.

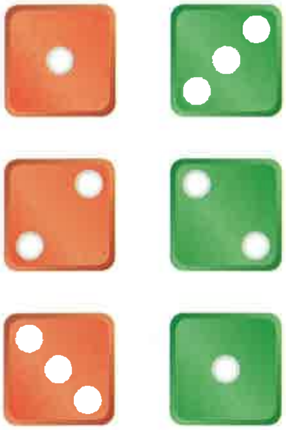


FIGURE 9.2 A sum of 4 on a roll of two dice. (Example 1d)

Sample Spaces and Probability Functions

Most people have an intuitive sense of probability. Unfortunately, this sense is not often based on a foundation of mathematical principles, so people become victims of scams, misleading statistics, and false advertising. In this lesson, we want to build on your intuitive sense of probability and give it a mathematical foundation.

EXAMPLE 1 Testing your intuition about probability

Find the probability of each of the following events.

(a) Tossing a head on one toss of a fair coin.

(b) Tossing two heads in a row on two tosses of a fair coin.

(c) Drawing a queen from a standard deck of 52 cards.

(d) Rolling a sum of 4 on a single roll of two fair dice.

(e) Guessing all 6 numbers in a state lottery that requires you to pick 6 numbers between 1 and 46, inclusive.

SOLUTION

(a) There are two equally likely outcomes: {T, H}. The probability is 1/2.

(b) There are four equally likely outcomes: {TT, TH, HT, HH}. The probability is 1/4.

(c) There are 52 equally likely outcomes, 4 of which are queens. The probability is 4/52, or 1/13.

(d) By the Multiplication Principle of Counting (Section 9.1), there are $6 \times 6 = 36$ equally likely outcomes. Of these, three {(1, 3), (3, 1), (2, 2)} yield a sum of 4 (Figure 9.2). The probability is 3/36, or 1/12.

(e) There are ${}_{46}C_6 = 9{,}366{,}819$ equally likely ways that 6 numbers can be chosen from 46 numbers without regard to order. Only one of these choices wins the lottery. The probability is $1/9{,}366{,}819 \approx 0.00000010676$.

Now try Exercise 5.

Is Probability Just For Games?

Probability theory got its start in letters between Blaise Pascal (1623–1662) and Pierre de Fermat (1601–1665) concerning games of chance, but it has come a long way since then. Modern mathematicians like David Blackwell (1919), the first African-American to receive a fellowship to the Institute for Advanced Study at Princeton, have greatly extended both the theory and the applications of probability, especially in the areas of statistics, quantum physics, and information theory. Moreover, the work of John Von Neumann (1903–1957) has led to a separate branch of modern discrete mathematics that really is about games, called Game Theory.

Notice that in each of these cases we first counted the number of possible outcomes of the experiment in question. The set of all possible outcomes of an experiment is the **sample space** of the experiment. An **event** is a subset of the sample space. Each of our sample spaces consisted of a finite number of **equally likely outcomes**, which enabled us to find the probability of an event by counting.

Probability of an Event (Equally Likely Outcomes)

If E is an event in a finite, nonempty sample space S of equally likely outcomes, then the **probability** of the event E is

$$P(E) = \frac{\text{the number of outcomes in } E}{\text{the number of outcomes in } S}.$$

The hypothesis of equally likely outcomes is critical here. Many people guess wrongly on the probability in Example 1d because they figure that there are 11 possible outcomes for the sum on two fair dice: {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} and that 4 is one of them. (That reasoning is correct so far.) The reason that 1/11 is not the probability of rolling a sum of 4 is that all those sums are *not equally likely*.

On the other hand, we can *assign* probabilities to the 11 outcomes in this smaller sample space in a way that is consistent with the number of ways each total can occur. The table below shows a **probability distribution**, in which each outcome is assigned a unique probability by a *probability function*.

Outcome	Probability
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

We see that the outcomes are not equally likely, but we can find the probabilities of events by adding up the probabilities of the outcomes in the event, as in the following example.

EXAMPLE 2 Rolling the dice

Find the probability of rolling a sum divisible by 3 on a single roll of two fair dice.

SOLUTION The event E consists of the outcomes $\{3, 6, 9, 12\}$. To get the probability of E we add up the probabilities of the outcomes in E (see the table of the probability distribution):

$$P(E) = \frac{2}{36} + \frac{5}{36} + \frac{4}{36} + \frac{1}{36} = \frac{12}{36} = \frac{1}{3}.$$

Now try Exercise 7.

Notice that this method would also have worked just fine with our 36-outcome sample space, in which every outcome has probability 1/36. In general, it is easier to work with sample spaces of equally likely events because it is not necessary to write out the probability distribution. When outcomes do have unequal probabilities, we need to know what probabilities to assign to the outcomes.

Not every function that assigns numbers to outcomes qualifies as a probability function.

EMPTY SET

A set with no elements is the *empty set*, denoted by $\varnothing$.

Definition Probability Function

A **probability function** is a function P that assigns a real number to each outcome in a sample space S subject to the following conditions:

1. $0 \leq P(O) \leq 1$ for every outcome O;
2. the sum of the probabilities of all outcomes in S is 1;
3. $P(\varnothing) = 0$.

The probability of any event can then be defined in terms of the probability function.

Probability of an Event (Outcomes not Equally Likely)

Let S be a finite, nonempty sample space in which every outcome has a probability assigned to it by a probability function P. If E is any event in S, the **probability** of the event E is the sum of the probabilities of all the outcomes contained in E.

EXAMPLE 3 Testing a probability function

Is it possible to weight a standard 6-sided die in such a way that the probability of rolling each number n is exactly $1/(n^2 + 1)$?

SOLUTION The probability distribution would look like this:

Outcome	Probability
1	1/2
2	1/5
3	1/10
4	1/17
5	1/26
6	1/37

This is not a valid probability function, because $1/2 + 1/5 + 1/10 + 1/17 + 1/26 + 1/37 \neq 1$.

Now try Exercise 9a.

Determining Probabilities

It is not always easy to determine probabilities, but the arithmetic involved is fairly simple. It usually comes down to multiplication, addition, and (most importantly) counting. Here is the strategy we will follow:

Strategy for Determining Probabilities

1. Determine the sample space of all possible outcomes. When possible, choose outcomes that are equally likely.
2. If the sample space has equally likely outcomes, the probability of an event E is determined by counting:

$$P(E) = \frac{\text{the number of outcomes in } E}{\text{the number of outcomes in } S}.$$

3. If the sample space does not have equally likely outcomes, determine the probability function. (This is not always easy to do.) Check to be sure that the conditions of a probability function are satisfied. Then the probability of an event E is determined by adding up the probabilities of all the outcomes contained in E.

EXAMPLE 4 Choosing chocolates, sample space I

Sal opens a box of a dozen chocolate cremes and generously offers two of them to Val. Val likes vanilla cremes the best, but all the chocolates look alike on the outside. If four of the twelve cremes are vanilla, what is the probability that both of Val's picks turn out to be vanilla?

SOLUTION The experiment in question is the selection of two chocolates, without regard to order, from a box of 12. There are ${}_{12}C_2 = 66$ outcomes of this experiment, and all of them are equally likely. We can therefore determine the probability by counting.

The event E consists of all possible pairs of 2 vanilla cremes that can be chosen, without regard to order, from 4 vanilla cremes available. There are ${}_4C_2 = 6$ ways to form such pairs.

Therefore, $P(E) = 6/66 = 1/11$.

Now try Exercise 25.

Many probability problems require that we think of events happening in succession, often with the occurrence of one event affecting the probability of the occurrence of another event. In these cases, we use a law of probability called the Multiplication Principle of Probability.

Multiplication Principle of Probability

Suppose an event A has probability p_1 and an event B has probability p_2 under the assumption that A occurs. Then the probability that both A and B occur is $p_1 p_2$.

If the events A and B are **independent**, we can omit the phrase "under the assumption that A occurs," since that assumption would not matter.

As an example of this principle at work, we will solve the *same problem* as that posed in Example 4, this time using a sample space that appears at first to be simpler, but which consists of events that are not equally likely.

EXAMPLE 5 Choosing chocolates, sample space II

Sal opens a box of a dozen chocolate cremes and generously offers two of them to Val. Val likes vanilla cremes the best, but all the chocolates look alike on the outside. If four of the twelve cremes are vanilla, what is the probability that both of Val's picks turn out to be vanilla?

SOLUTION As far as Val is concerned, there are two kinds of chocolate cremes: vanilla (V) and unvanilla (U). When choosing two chocolates, there are four possible outcomes: VV, VU, UV, and UU. We need to determine the probability of the outcome VV.

Notice that these four outcomes are *not equally likely*! There are twice as many U chocolates as V chocolates. So we need to consider the distribution of probabilities, and we may as well begin with $P(VV)$, as that is the probability we seek.

The probability of picking a vanilla creme on the first draw is $4/12$. The probability of picking a vanilla creme on the second draw, *under the assumption that a vanilla creme was drawn on the first*, is $3/11$. By the Multiplication Principle, the probability of drawing a vanilla creme on both draws is

$$\frac{4}{12} \cdot \frac{3}{11} = \frac{1}{11}.$$

ORDERED OR UNORDERED?

Notice that in Example 4 we had a sample space in which order was disregarded, whereas in Example 5 we have a sample space in which order matters. (For example, *UV* and *VU* are distinct outcomes.) The order matters in Example 5 because we are considering the probabilities of two events (first draw, second draw), one of which affects the other. In Example 4, we are simply counting unordered combinations.

Since this is the probability we are looking for, we do not need to compute the probabilities of the other outcomes. However, you should verify that the other probabilities would be:

$$P(VU) = \frac{4}{12} \cdot \frac{8}{11} = \frac{8}{33}$$
$$P(UV) = \frac{8}{12} \cdot \frac{4}{11} = \frac{8}{33}$$
$$P(UU) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}$$

Notice that $P(VV) + P(VU) + P(UV) + P(UU) = (1/11) + (8/33) + (8/33) + (14/33) = 1$, so the probability function checks out.

Now try Exercise 33.

JOHN VENN

John Venn (1834–1923) was an English logician and clergyman, just like his contemporary, Charles L. Dodgson. Although both men used overlapping circles to illustrate their logical syllogisms, it is Venn whose name lives on in connection with these diagrams. Dodgson's name barely lives on at all, and yet he is far the more famous of the two: under the pen name Lewis Carroll, he wrote *Alice's Adventures in Wonderland* and *Through the Looking Glass*.

Venn Diagrams and Tree Diagrams

We have seen many instances in which geometric models help us to understand algebraic models more easily, and probability theory is yet another setting in which this is true. **Venn diagrams**, associated mainly with the world of set theory, are good for visualizing relationships among events within sample spaces. **Tree diagrams**, which we first met in Section 9.1 as a way to visualize the Multiplication Principle of Counting, are good for visualizing the Multiplication Principle of Probability.

EXAMPLE 6 Using a Venn diagram

In a large high school, 54% of the students are girls and 62% of the students play sports. Half of the girls at the school play sports.

(a) What percentage of the students who play sports are boys?

(b) If a student is chosen at random, what is the probability that it is a boy who does not play sports?

SOLUTION To organize the categories, we draw a large rectangle to represent the sample space (all students at the school) and two overlapping regions to represent "girls" and "sports" (Figure 9.3). We fill in the percentages (Figure 9.4) using the following logic:

- The overlapping (green) region contains half the girls, or (0.5)(54%) = 27% of the students.
- The yellow region (the rest of the girls) then contains (54 – 27)% = 27% of the students.
- The blue region (the rest of the sports players) then contains (62 – 27)% = 35% of the students.
- The white region (the rest of the students) then contains (100 – 89)% = 11% of the students. These are boys who do not play sports.

We can now answer the two questions by looking at the Venn diagram.

(a) We see from the diagram that the ratio of *boys* who play sports to *all students* who play sports is $\frac{0.35}{0.62}$, which is about 56.45%.

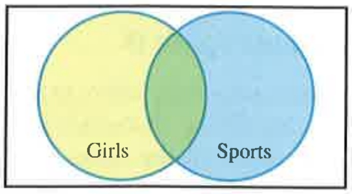


FIGURE 9.3 A Venn diagram for Example 6. The overlapping region common to both circles represents "girls who play sports." The region outside both circles (but inside the rectangle) represents "boys who do not play sports."

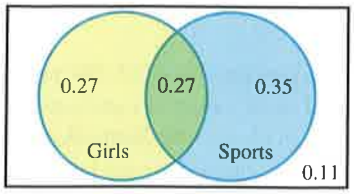


FIGURE 9.4 A Venn diagram for Example 6 with the probabilities filled in.

(b) We see that 11% of the students are boys who do not play sports, so 0.11 is the probability. Now try Exercises 27a–d.

EXAMPLE 7 Using a tree diagram

Two identical cookie jars are on a counter. Jar A contains 2 chocolate chip and 2 peanut butter cookies, while jar B contains 1 chocolate chip cookie. We select a cookie at random. What is the probability that it is a chocolate chip cookie?

SOLUTION It is tempting to say 3/5, since there are 5 cookies in all, 3 of which are chocolate chip. Indeed, this would be the answer if all the cookies were in the same jar. However, the fact that they are in different jars means that the 5 cookies are *not equally likely outcomes*. That lone chocolate chip cookie in jar B has a much better chance of being chosen than any of the cookies in jar A. We need to think of this as a two-step experiment: first pick a jar, then pick a cookie from that jar.

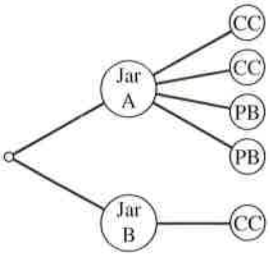


FIGURE 9.5 A tree diagram for Example 7.

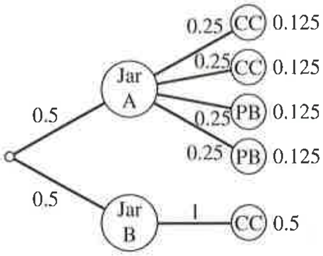


FIGURE 9.6 The tree diagram for Example 7 with the probabilities filled in. Notice that the five cookies are not equally likely to be drawn. Notice also that the probabilities of the five cookies do add up to 1.

Figure 9.5 gives a visualization of the two-step process. In Figure 9.6, we have filled in the probabilities along each branch, first of picking the jar, then of picking the cookie. The probability at the *end* of each branch is obtained by multiplying the probabilities from the root to the branch. (This is the Multiplication Principle.) Notice that the probabilities of the 5 cookies (as predicted) are not equal.

The event "chocolate chip" is a set containing three outcomes. We add their probabilities together to get the correct probability:

$$P(\text{chocolate chip}) = 0.125 + 0.125 + 0.5 = 0.75.$$

Now try Exercise 29.

Conditional Probability

The probability of drawing a chocolate chip cookie in Example 7 is an example of **conditional probability**, since the "cookie" probability is **dependent** on the "jar" outcome. A convenient symbol to use with conditional probability is $P(A|B)$, pronounced "P of A given B," meaning "the probability of the event A, given that event B occurs." In the cookie jars of Example 7,

$$P(\text{chocolate chip}|\text{jar } A) = \frac{2}{4} \quad \text{and} \quad P(\text{chocolate chip}|\text{jar } B) = 1.$$

(In the tree diagram, these are the probabilities along the *branches* that come out of the two jars, not the probabilities at the *ends* of the branches.)

The Multiplication Principle of Probability can be stated succinctly with this notation as follows:

$$P(A \text{ and } B) = P(A) \cdot P(B|A).$$

This is how we found the numbers at the ends of the branches in Figure 9.6.

As our final example of a probability problem, we will show how to use this formula in a different but equivalent form, sometimes called the **conditional probability formula**:

Conditional Probability Formula

If the event B depends on the event A, then $P(B \mid A) = \dfrac{P(A \text{ and } B)}{P(A)}$.

EXAMPLE 8 Using the conditional probability formula

Suppose we have drawn a cookie at random from one of the jars described in Example 7. Given that it is chocolate chip, what is the probability that it came from jar A?

SOLUTION By the formula,

$$P(\text{jar } A \mid \text{chocolate chip}) = \frac{P(\text{jar } A \text{ and chocolate chip})}{P(\text{chocolate chip})}$$

$$= \frac{(1/2)(2/4)}{0.75} = \frac{0.25}{0.75} = \frac{1}{3}$$

Now try Exercise 31.

EXPLORATION 1 Testing Positive for HIV

As of the year 2000, the probability of an adult (age 15–49) in North America having HIV/AIDS was 0.0061 (Source: *UNAIDS/WHO*). The ELISA test is used to detect the virus antibody in blood. If the antibody is present, the test reports positive with probability 0.997 and negative with probability 0.003. If the antibody is not present, the test reports positive with probability 0.015 and negative with probability 0.985.

1. Draw a tree diagram with branches to nodes "antibody present" and "antibody absent" branching from the root. Fill in the probabilities for North American adults (age 15–49) along the branches. (Note that these two probabilities must add up to 1.)
2. From the node at the end of each of the two branches, draw branches to "positive" and "negative." Fill in the probabilities along the branches.
3. Use the Multiplication Principle to fill in the probabilities at the ends of the four branches. Check to see that they add up to 1.
4. Find the probability of a positive test result. (Note that this event consists of two outcomes).
5. Use the conditional probability formula to find the probability that a person with a positive test result actually *has* the antibody, i.e., $P(\text{antibody present} \mid \text{positive})$.

You might be surprised that the answer to part 5 is so low, but it should be compared with the probability of the antibody being present *before* seeing the positive test result, which was 0.0061. Nonetheless, that is why a positive ELISA test is followed by further testing before a diagnosis of HIV/AIDS is made. This is the case with many diagnostic tests.

Binomial Distributions

We noted in our "Strategy for Determining Probabilities" that it is not always easy to determine a probability distribution for a sample space with unequal probabilities. An interesting exception for those who have studied the Binomial Theorem (Section 9.2) is the binomial distribution.

EXAMPLE 9 Repeating a simple experiment

We roll a fair die four times. Find the probability that we roll:

(a) all 3's. **(b)** no 3's. **(c)** exactly two 3's.

SOLUTION

(a) We have a probability $1/6$ of rolling a 3 each time. By the Multiplication Principle, the probability of rolling a 3 all four times is $(1/6)^4 \approx 0.00077$.

(b) There is a probability $5/6$ of rolling something other than 3 each time. By the Multiplication Principle, the probability of rolling a non-3 all four times is $(5/6)^4 \approx 0.48225$.

(c) The probability of rolling two 3's followed by two non-3's (again by the Multiplication Principle) is $(1/6)^2(5/6)^2 \approx 0.01929$. However, that is not the only outcome we must consider. In fact, the two 3's could occur *anywhere* among the four rolls, in exactly $\binom{4}{2} = 6$ ways. That gives us 6 outcomes, each with probability $(1/6)^2(5/6)^2$. The probability of the event "exactly two 3's" is therefore $\binom{4}{2}(1/6)^2(5/6)^2 \approx 0.11574$.

Now try Exercise 47.

Did the form of those answers look a little familiar? Watch what they look like when we let $p = 1/6$ and $q = 5/6$:

$$P(\text{four 3's}) = p^4$$

$$P(\text{no 3's}) = q^4$$

$$P(\text{two 3's}) = \binom{4}{2} p^2q^2$$

You can probably recognize these as three of the terms in the expansion of $(p + q)^4$. This is no coincidence. In fact, the terms in the expansion

$$(p + q)^4 = p^4 + 4p^3q^1 + 6p^2q^2 + 4p^1q^3 + q^4$$

give the exact probabilities of 4, 3, 2, 1, and 0 threes (respectively) when we toss a fair die four times! That is why this is called a *binomial probability distribution*. The general theorem follows.

BINOMIAL PROBABILITIES ON A CALCULATOR

Your calculator might be programmed to find values for the binomial probability distribution function (binompdf). The solutions to Example 10 in one calculator syntax, for example, could be obtained by:

(a) binompdf(20, .9, 20) (20 repetitions, 0.9 probability, 20 successes)

(b) binompdf(20, .9, 18) (20 repetitions, 0.9 probability, 18 successes)

(c) 1 − binomcdf(20, .9, 17) (1 minus the *cumulative* probability of 17 or fewer successes)

Check your owner's manual for more information.

Theorem **Binomial Distribution**

Suppose an experiment consists of n independent repetitions of an experiment with two outcomes, called "success" and "failure." Let $P(\text{success}) = p$ and $P(\text{failure}) = q$. (Note that $q = 1 - p$.)

Then the terms in the binomial expansion of $(p + q)^n$ give the respective probabilities of exactly $n, n - 1, \ldots, 2, 1, 0$ successes. The distribution is shown below:

Number of successes out of n independent repetitions	Probability
n	p^n
$n - 1$	$\binom{n}{n-1}p^{n-1}q$
⋮	⋮
r	$\binom{n}{r}p^r q^{n-r}$
⋮	⋮
1	$\binom{n}{1}pq^{n-1}$
0	q^n

EXAMPLE 10 Shooting free throws

Suppose Michael makes 90% of his free throws. If he shoots 20 free throws, and if his chance of making each one is independent of the other shots (an assumption you might question in a game situation), what is the probability that he makes

(a) all 20?

(b) exactly 18?

(c) at least 18?

SOLUTION We could get the probabilities of all possible outcomes by expanding $(0.9 + 0.1)^{20}$, but that is not necessary in order to answer these three questions. We just need to compute three specific terms.

(a) $P(20 \text{ successes}) = (0.9)^{20} \approx 0.12158$

(b) $P(18 \text{ successes}) = \binom{20}{18}(0.9)^{18}(0.1)^2 \approx 0.28518$

(c) $P(\text{at least } 18 \text{ successes}) = P(18) + P(19) + P(20)$

$$= \binom{20}{18}(0.9)^{18}(0.1)^2 + \binom{20}{19}(0.9)^{19}(0.1) + (0.9)^{20}$$

$$\approx 0.6769$$

Now try Exercise 49.

QUICK REVIEW 9.3

(Prerequisite skill Section 9.1)

In Exercises 1–8, tell how many outcomes are possible for the experiment.

1. A single coin is tossed.

2. A single 6-sided die is rolled.

3. Three different coins are tossed.

4. Three different 6-sided dice are rolled.

5. Five different cards are drawn from a standard deck of 52.

6. Two chips are drawn simultaneously without replacement from a jar containing 10 chips.

7. Five people are lined up for a photograph.

8. Three-digit numbers are formed from the numbers $\{1, 2, 3, 4, 5\}$ without repetition.

In Exercises 9 and 10, evaluate the expression by pencil and paper. Verify your answer with a calculator.

9. $\frac{{}_5C_3}{{}_{10}C_3}$

10. $\frac{{}_5C_2}{{}_{10}C_2}$

SECTION 9.3 EXERCISES

In Exercises 1–8, a red die and a green die have been rolled. What is the probability of the event?

1. The sum is 9.

2. The sum is even.

3. The number on the red die is greater than the number on the green die.

4. The sum is less than 10.

5. Both numbers are odd.

6. Both numbers are even.

7. The sum is prime.

8. The sum is 7 or 11.

9. **Writing to Learn** Alrik's gerbil cage has four compartments, A, B, C, and D. After careful observation, he estimates the proportion of time the gerbil spends in each compartment and constructs the table below.

Compartment	A	B	C	D
Proportion	0.25	0.20	0.35	0.30

(a) Is this a valid probability function? Explain.

(b) Is there a problem with Alrik's reasoning? Explain.

10. (Continuation of Exercise 9) Suppose Alrik determines that his gerbil spends time in the four compartments A, B, C, and D in the ratio 4:3:2:1. What proportions should he fill in the table above? Is this a valid probability function?

The maker of a popular chocolate candy that is covered in a thin colored shell has released information about the overall color proportions in its production of the candy, which is summarized in the following table.

Color	Brown	Red	Yellow	Green	Orange	Tan
Proportion	0.3	0.2	0.2	0.1	0.1	0.1

In Exercises 11–16, a single candy of this type is selected at random from a newly-opened bag. What is the probability that the candy has the given color(s)?

11. Brown or tan

12. Red, green, or orange

13. Red

14. Not red

15. Neither orange nor yellow

16. Neither brown nor tan

A peanut version of the same candy has all the same colors except tan. The proportions of the peanut version are given in the following table.

Color	Brown	Red	Yellow	Green	Orange
Proportion	0.3	0.2	0.2	0.2	0.1

In Exercises 17–22, a candy of this type is selected at random from each of two newly-opened bags. What is the probability that the two candies have the given color(s)?

17. Both are brown.

18. Both are orange.

19. One is red, and the other is green.

20. The first is brown, and the second is yellow.

21. Neither is yellow.

22. The first is not red, and the second is not orange.

Exercises 23–32 concern a version of the card game "bid Euchre" that uses a pack of 24 cards, consisting of ace, king, queen, jack, 10, and 9 in each of the four suits (spades, hearts, diamonds, and clubs). In bid Euchre, a hand consists of 6 cards. Find the probability of each event.

23. **Euchre** A hand is all spades.

24. **Euchre** All six cards are from the same suit.

25. **Euchre** A hand includes all four aces.

26. **Euchre** A hand includes two jacks of the same color (called the right and left bower).

27. **Using Venn Diagrams** A and B are events in a sample space S such that $P(A) = 0.6$, $P(B) = 0.5$, and $P(A \text{ and } B) = 0.3$.

(a) Draw a Venn diagram showing the overlapping sets A and B and fill in the probabilities of the four regions formed.

(b) Find the probability that A occurs but B does not.

(c) Find the probability that B occurs but A does not.

(d) Find the probability that neither A nor B occurs.

(e) Are events A and B independent? (That is, does $P(A \mid B) = P(A)$?)

28. **Using Venn Diagrams** A and B are events in a sample space S such that $P(A) = 0.7$, $P(B) = 0.4$, and $P(A \text{ and } B) = 0.2$.

(a) Draw a Venn diagram showing the overlapping sets A and B and fill in the probabilities of the four regions formed.

(b) Find the probability that A occurs but B does not.

(c) Find the probability that B occurs but A does not.

(d) Find the probability that neither A nor B occurs.

(e) Are events A and B independent? (That is, does $P(A \mid B) = P(A)$?)

In Exercises 29 and 30, it will help to draw a tree diagram.

29. **Piano Lessons** If it rains tomorrow, the probability is 0.8 that John will practice his piano lesson. If it does not rain tomorrow, there is only a 0.4 chance that John will practice. Suppose that the chance of rain tomorrow is 60%. What is the probability that John will practice his piano lesson?

30. **Predicting Cafeteria Food** If the school cafeteria serves meat loaf, there is a 70% chance that they will serve peas. If they do not serve meat loaf, there is a 30% chance that they will serve peas anyway. The students know that meat loaf will be served exactly once during the 5-day week, but they do not know which day. If tomorrow is Monday, what is the probability that

(a) the cafeteria serves meat loaf?

(b) the cafeteria serves meat loaf and peas?

(c) the cafeteria serves peas?

31. **Conditional Probability** There are two precalculus sections at West High School. Mr. Abel's class has 12 girls and 8 boys, while Mr. Bonitz's class has 10 girls and 15 boys. If a West High precalculus student chosen at random happens to be a girl, what is the probability she is from Mr. Abel's class? [*Hint*: The answer is not 12/22.]

32. **Group Activity** **Conditional Probability** Two boxes are on the table. One box contains a normal coin and a two-headed coin; the other box contains three normal coins. A friend reaches into a box, removes a coin, and shows you one side: a head. What is the probability that it came from the box with the two-headed coin?

33. **Renting Cars** Floppy Jalopy Rent-a-Car has 25 cars available for rental—20 big bombs and 5 midsize cars. If two cars are selected at random, what is the probability that both are big bombs?

34. **Defective Calculators** Dull Calculators, Inc., knows that a unit coming off an assembly line has a probability of 0.037 of being defective. If four units are selected at random during the course of a workday, what is the probability that none of the units are defective?

35. **Causes of Death** The government designates a single cause for each death in the United States. The resulting data indicate that 45% of deaths are due to heart and other cardiovascular disease and 22% are due to cancer.

(a) What is the probability that the death of a randomly selected person will be due to cardiovascular disease or cancer?

(b) What is the probability that the death will be due to some other cause?

36. **Yahtzee** In the game of *Yahtzee*, on the first roll five dice are tossed simultaneously. What is the probability of rolling five of a kind (which is Yahtzee!) on the first roll?

37. **Writing to Learn** Explain why the following statement cannot be true. The probabilities that a computer salesperson will sell zero, one, two, or three computers in any one day are 0.12, 0.45, 0.38, and 0.15, respectively.

38. **HIV Testing** A particular test for HIV, the virus that causes AIDS, is 0.7% likely to produce a false positive result—a result indicating that the human subject has HIV when in fact the person is not carrying the virus. If 60 individuals who are HIV-negative are tested, what is the probability of obtaining at least one false result?

39. **Graduate School Survey** The Earmuff Junction College Alumni Office surveys selected members of the class of 2000. Of the 254 who graduated that year, 172 were women, 124 of whom went on to graduate school. Of the male graduates, 58 went on to graduate school. What is the probability of the given event?

(a) The graduate is a woman.

(b) The graduate went on to graduate school.

(c) The graduate was a woman who went on to graduate school.

40. **Indiana Jones and the Final Exam** Professor Indiana Jones gives his class a list of 20 study questions, from which he will select 8 to be answered on the final exam. If a given student knows how to answer 14 of the questions, what is the probability that the student will be able to answer the given number of questions correctly?

(a) All 8 questions

(b) Exactly 5 questions

(c) At least 6 questions

41. **Graduation Requirement** To complete the kinesiology requirement at Palpitation Tech you must pass two classes chosen from aerobics, aquatics, defense arts, gymnastics, racket sports, recreational activities, rhythmic activities, soccer, and volleyball. If you decide to choose your two classes at random by drawing two class names from a box, what is the probability that you will take racket sports and rhythmic activities?

42. **Writing to Learn** During July in Gunnison, Colorado, the probability of at least 1 hour a day of sunshine is 0.78, the probability of at least 30 minutes of rain is 0.44, and the probability that it will be cloudy all day is 0.22. Write a paragraph explaining whether this statement could be true.

In Exercises 43–50, ten dimes dated 1990 through 1999 are tossed. Find the probability of each event.

43. **Tossing Ten Dimes** Heads on the 1990 dime only

44. **Tossing Ten Dimes** Heads on the 1991 and 1996 dimes only

45. **Tossing Ten Dimes** Heads on all 10 dimes

46. **Tossing Ten Dimes** Heads on all but one dime

47. **Tossing Ten Dimes** Exactly two heads

48. **Tossing Ten Dimes** Exactly three heads

49. **Tossing Ten Dimes** At least one head

50. **Tossing Ten Dimes** At least two heads

Standardized Test Questions

51. **True or False** A sample space consists of equally likely events. Justify your answer.

52. **True or False** The probability of an event can be greater than 1. Justify your answer.

Evaluate Exercises 53–56 without using a calculator.

53. The probability of rolling a total of 5 on a pair of fair dice is

(a) $\frac{1}{4}$.

(b) $\frac{1}{5}$.

(c) $\frac{1}{6}$.

(d) $\frac{1}{9}$.

(e) $\frac{1}{11}$.

54. Which of the following numbers could not be the probability of an event?

(a) 0

(b) 0.95

(c) $\frac{\sqrt{3}}{4}$

(d) $\frac{3}{\pi}$

(e) $\frac{\pi}{2}$

55. If A and B are independent events, then $P(A \mid B) =$

(a) $P(A)$.

(b) $P(B)$.

(c) $P(B \mid A)$.

(d) $P(A) \cdot P(B)$.

(e) $P(A) + P(B)$.

56. A fair coin is tossed three times in succession. What is the probability that exactly one of the coins shows heads?

(a) $\frac{1}{8}$

(b) $\frac{1}{3}$

(c) $\frac{3}{8}$

(d) $\frac{1}{2}$

(e) $\frac{2}{3}$

Explorations

57. Empirical Probability In real applications, it is often necessary to approximate the probabilities of the various outcomes of an experiment by performing the experiment a large number of times and recording the results. Barney's Bread Basket offers five different kinds of bagels. Barney records the sales of the first 500 bagels in a given week in the table shown below:

Type of Bagel	Number Sold
Plain	185
Onion	60
Rye	55
Cinnamon Raisin	125
Sourdough	75

(a) Use the observed sales number to approximate the probability that a random customer buys a plain bagel. Do the same for each other bagel type and make a table showing the approximate probability distribution.

(b) Assuming independence of the events, find the probability that three customers in a row all order plain bagels.

(c) Writing to Learn Do you think it is reasonable to assume that the orders of three consecutive customers actually are independent? Explain.

58. Straight Poker In the original version of poker known as "straight" poker, a 5-card hand is dealt from a standard deck of 52 cards. What is the probability of the given event?

(a) A hand will contain at least one king.

(b) A hand will be a "full house" (any three of one kind and a pair of another kind).

59. Married Students Suppose that 23% of all college students are married. Answer the following questions for a random sample of eight college students.

(a) How many would you expect to be married?

(b) Would you regard it as unusual if the sample contained five married students?

(c) What is the probability that five or more of the eight students are married?

60. Group Activity Investigating an Athletic Program A university widely known for its track and field program claims that 75% of its track athletes get degrees. A journalist investigates what happened to the 32 athletes who began the program over a 6-year period that ended 7 years ago. Of these athletes, 17 have graduated and the remaining 15 are no longer attending any college. If the university's claim is true, the number of athletes who graduates among the 32 examined should have been governed by binomial probability with $p = 0.75$.

(a) What is the probability that exactly 17 athletes should have graduated?

(b) What is the probability that 17 or fewer athletes should have graduated?

(c) If you were the journalist, what would you say in your story on the investigation?

Extending the Ideas

61. Expected Value If the outcomes of an experiment are given numerical values (such as the total on a roll of two dice, or the payoff on a lottery ticket), we define the **expected value** to be the sum of all the numerical values times their respective probabilities.

For example, suppose we roll a fair die. If we roll a multiple of 3, we win \$3; otherwise we lose \$1. The probabilities of the two possible payoffs are shown in the table below:

Value	Probability
+3	2/6
−1	4/6

The expected value is

$3 \times (2/6) + (-1) \times (4/6) = (6/6) - (4/6) = 1/3.$

We interpret this to mean that we would win an average of 1/3 dollar per game in the long run.

(a) A game is called *fair* if the expected value of the payoff is zero. Assuming that we still win \$3 for a multiple of 3, what should we pay for any other outcome in order to make the game fair?

(b) Suppose we roll *two* fair dice and look at the total under the original rules. That is, we win \$3 for rolling a multiple of 3 and lose \$1 otherwise. What is the expected value of this game?

62. Expected Value (Continuation of Exercise 61) Gladys has a personal rule never to enter the lottery (picking 6 numbers from 1 to 46) until the payoff reaches 4 million dollars. When it does reach 4 million, she always buys ten different \$1 tickets.

(a) Assume that the payoff for a winning ticket is 4 million dollars. What is the probability that Gladys holds a winning ticket? (Refer to Example 1 of this section for the probability of any ticket winning.)

(b) Fill in the probability distribution for Gladys's possible payoffs in the table below. (Note that we subtract \$10 from the \$4 million, since Gladys has to pay for her tickets even if she wins.)

Value	Probability
−10	
+3,999,990	

(c) Find the expected value of the game for Gladys.

(d) Writing to Learn In terms of the answer in part (b), explain to Gladys the long-term implications of her strategy.

9.4 SEQUENCES AND SERIES

What you'll learn about

- Infinite Sequences
- Arithmetic and Geometric Sequences
- Sequences and Graphing Calculators
- Summation Notation
- Sums of Arithmetic and Geometric Sequences
- Infinite Series
- Convergence of Geometric Series

. . . and why

Infinite sequences and series are important bridges to calculus.

Infinite Sequences

One of the most natural ways to study patterns in mathematics is to look at an ordered progression of numbers, called a **sequence**. Here are some examples of sequences:

1. 5, 10, 15, 20, 25

2. $2, 4, 8, 16, 32, \ldots, 2^k, \ldots$

3. $\left\{\dfrac{1}{k}: k = 1, 2, 3, \ldots\right\}$

4. $\{a_1, a_2, a_3, \ldots, a_k, \ldots\}$, which is sometimes abbreviated $\{a_k\}$

The first of these is a **finite sequence**, while the other three are **infinite sequences**. Notice that in (2) and (3) we were able to define a rule that gives the kth number in the sequence (called the **kth term**) as a function of k. In (4) we do not have a rule, but notice how we can use subscript notation (a_k) to identify the kth term of a "general" infinite sequence. In this sense, an infinite sequence can be thought of as a *function* that assigns a unique number (a_k) to each natural number k.

EXAMPLE 1 Defining a sequence explicitly

Find the first 6 terms and the 100th term of the sequence $\{a_k\}$ in which $a_k = k^2 - 1$.

SOLUTION Since we know the kth term *explicitly* as a function of k, we need only to evaluate the function to find the required terms:

$$a_1 = 1^2 - 1 = 0,\ a_2 = 3,\ a_3 = 8,\ a_4 = 15,\ a_5 = 24,\ a_6 = 35, \quad \text{and}$$

$$a_{100} = 100^2 - 1 = 9999.$$

Now try Exercise 1.

Explicit formulas are the easiest to work with, but there are other ways to define sequences. For example, we can specify values for the first term (or terms) of a sequence, then define each of the following terms **recursively** by a formula relating it to previous terms. Example 2 shows how this is done.

AGREEMENT ON SEQUENCES

Since we will be dealing primarily with infinite sequences in this book, the word "sequence" will mean an infinite sequence unless otherwise specified.

EXAMPLE 2 Defining a sequence recursively

Find the first 6 terms and the 100th term for the sequence defined recursively by the conditions:

$$b_1 = 3$$

$$b_n = b_{n-1} + 2 \text{ for all } n > 1.$$

SOLUTION We proceed one term at a time, starting with $b_1 = 3$ and obtaining each succeeding term by adding 2 to the term just before it:

$$b_1 = 3$$
$$b_2 = \overset{2}{b_1} + 2 = 5$$
$$b_3 = b_2 + 2 = 7$$

etc.

Eventually it becomes apparent that we are building the sequence of odd natural numbers beginning with 3:

$$\{3, 5, 7, 9, \ldots\}.$$

The 100th term is 99 terms beyond the first, which means that we can get there quickly by adding 99 2's to the number 3:

$$b_{100} = 3 + 99 \times 2 = 201.$$

Now try Exercise 5.

Arithmetic and Geometric Sequences

There are all kinds of rules by which we can construct sequences, but two particular types of sequences dominate in mathematical applications: those in which pairs of successive terms all have a common *difference* (**arithmetic** sequences), and those in which pairs of successive terms all have a common quotient, or *ratio* (**geometric** sequences). We will take a closer look at these in this section.

PRONUNCIATION TIP

The word "arithmetic" is probably more familiar to you as a noun, referring to the mathematics you studied in elementary school. In this word, the second syllable ("rith") is accented. When used as an adjective, the third syllable ("met") gets the accent. (For the sake of comparison, a similar shift of accent occurs when going from the noun "analysis" to the adjective "analytic.")

Definition Arithmetic Sequence

A sequence $\{a_n\}$ is an **arithmetic sequence** if it can be written in the form

$$\{a, a + d, a + 2d, \ldots, a + (n - 1)d, \ldots\} \text{ for some constant } d.$$

The number d is called the **common difference**.

Each term in an arithmetic sequence can be obtained recursively from its preceding term by adding d:

$$a_n = a_{n-1} + d \text{ (for all } n \geq 2).$$

EXAMPLE 3 Defining arithmetic sequences

For each of the following arithmetic sequences, find **(a)** the common difference, **(b)** the tenth term, **(c)** a recursive rule for the nth term, and **(d)** an explicit rule for the nth term.

(1) $-6, -2, 2, 6, 10, \ldots$

(2) $\ln 3, \ln 6, \ln 12, \ln 24, \ldots$

SOLUTION

(1) **(a)** The difference between successive terms is 4.

(b) $a_{10} = -6 + (10 - 1)(4) = 30$

(c) The sequence is defined recursively by $a_1 = -6$ and $a_n = a_{n-1} + 4$ for all $n \geq 2$.

(d) The sequence is defined explicitly by $a_n = -6 + (n - 1)(4) = 4n - 10$.

(2) **(a)** This sequence might not look arithmetic at first, but $\ln 6 - \ln 3 = \ln \frac{6}{3} = \ln 2$ (by a law of logarithms) and the difference between successive terms continues to be ln 2.

(b) $a_{10} = \ln 3 + (10 - 1)\ln 2 = \ln 3 + 9 \ln 2 = \ln(3 \cdot 2^9) = \ln 1536$

(c) The sequence is defined recursively by $a_1 = \ln 3$ and $a_n = a_{n-1} + \ln 2$ for all $n \geq 2$.

(d) The sequence is defined explicitly by $a_n = \ln 3 + (n - 1)\ln 2 = \ln (3 \cdot 2^{n-1})$.

Now try Exercise 11.

Definition Geometric Sequence

A sequence $\{a_n\}$ is a **geometric sequence** if it can be written in the form

$\{a, a \cdot r, a \cdot r^2, \ldots, a \cdot r^{n-1}, \ldots\}$ for some nonzero constant r.

The number r is called the **common ratio**.

Each term in a geometric sequence can be obtained recursively from its preceding term by multiplying by r:

$$a_n = a_{n-1} \cdot r \text{ (for all } n \geq 2).$$

EXAMPLE 4 Defining geometric sequences

For each of the following geometric sequences, find **(a)** the common ratio, **(b)** the tenth term, **(c)** a recursive rule for the *n*th term, and **(d)** an explicit rule for the *n*th term.

(1) 3, 6, 12, 24, 48, . . .

(2) $10^{-3}, 10^{-1}, 10^1, 10^3, 10^5, \ldots$

SOLUTION

(1) **(a)** The ratio between successive terms is 2.

(b) $a_{10} = 3 \cdot 2^{10-1} = 3 \cdot 2^9 = 1536$

(c) The sequence is defined recursively by $a_1 = 3$ and $a_n = 2a_{n-1}$ for $n \geq 2$.

(d) The sequence is defined explicitly by $a_n = 3 \cdot 2^{n-1}$.

(2) **(a)** Applying a law of exponents, $\frac{10^{-1}}{10^{-3}} = 10^{-1-(-3)} = 10^2$, and the ratio between successive terms continues to be 10^2.

(b) $a_{10} = 10^{-3} \cdot (10^2)^{10-1} = 10^{-3+18} = 10^{15}$

(c) The sequence is defined recursively by $a_1 = 10^{-3}$ and $a_n = 10^2 a_{n-1}$ for $n \geq 2$.

(d) The sequence is defined explicitly by $a_n = 10^{-3}(10^2)^{n-1} = 10^{-3+2n-2} = 10^{2n-5}$.

Now try Exercise 15.

EXAMPLE 5 Constructing sequences

The second and fifth terms of a sequence are 3 and 24, respectively. Find explicit and recursive formulas for the sequence if it is **(a)** arithmetic and **(b)** geometric.

SOLUTION

(a) If the sequence is arithmetic, then $a_2 = a + d = 3$ and $a_5 = a + 4d = 24$. Subtracting, we have

$$(a + 4d) - (a + d) = 24 - 3$$
$$3d = 21$$
$$d = 7$$

Then $a + d = 3$ implies $a = -4$.

The sequence is defined explicitly by $a_n = -4 + (n - 1) \cdot 7$, or $a_n = 7n - 11$.

The sequence is defined recursively by $a_1 = -4$ and $a = a_{n-1} + 7$ for $n \geq 2$.

(b) If the sequence is geometric, then $a_2 = a \cdot r^1 = 3$ and $a_5 = a \cdot r^4 = 24$. Dividing, we have

$$\frac{a \cdot r^4}{a \cdot r^1} = \frac{24}{3}$$
$$r^3 = 8$$
$$r = 2$$

Then $a \cdot r^1 = 3$ implies $a = 1.5$.

The sequence is defined explicitly by $a_n = 1.5(2)^{n-1}$, or $a_n = 3(2)^{n-2}$.

The sequence is defined recursively by $a_1 = 1.5$ and $a = 2 \cdot a_{n-1}$

Now try Exercise 19.

Sequences and Graphing Calculators

As with other kinds of functions, it helps to be able to represent a sequence geometrically with a graph. There are at least two ways to obtain a sequence graph on a graphing calculator. One way to graph explicitly defined sequences is as scatter plots of points of the form (k, a_k). A second way is to use the sequence graphing mode on a graphing calculator.

SEQUENCE GRAPHING

Most graphers enable you to graph in "sequence mode." Check your owner's manual to see how to use this mode.

EXAMPLE 6 Graphing a sequence defined explicitly

Produce on a graphing calculator a graph of the sequence $\{a_k\}$ in which $a_k = k^2 - 1$.

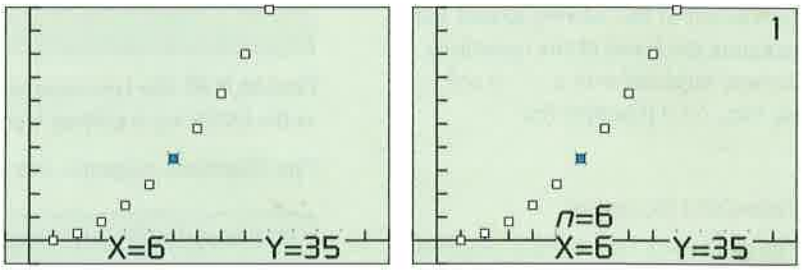
Method 1 (Scatter Plot)

The command seq(K, K, 1, 10) $\to$ L_1 puts the first 10 natural numbers in list L_1. (You could change the 10 if you wanted to graph more or fewer points.)

The command $L_1^2 - 1 \to L_2$ puts the corresponding terms of the sequence in list L_2. A scatter plot of L_1, L_2 produces the graph in Figure 9.7a.

Method 2 (Sequence Mode)

With your calculator in Sequence mode, enter the sequence $a_k = k^2 - 1$ in the Y = list as $u(n) = n^2 - 1$ with nMin = 1, nMax = 10, and $u(n\text{Min}) = 0$. (You could change the 10 if you wanted to graph more or fewer points.) Figure 9.7b shows the graph in the same window as Figure 9.7a.



[–1, 15] by [–10, 100]

(a)

[–1, 15] by [–10, 100]

(b)

FIGURE 9.7 The sequence $a_k = k^2 - 1$ graphed (a) as a scatter plot and (b) using the sequence graphing mode. Tracing along the points gives values of a_k for $k = 1, 2, 3, \ldots$ (Example 6)

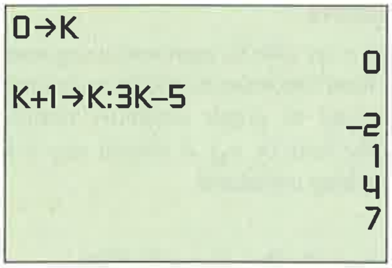
Now try Exercise 23.

EXAMPLE 7 Generating a sequence with a calculator

Using a graphing calculator, generate the specific terms of the following sequences:

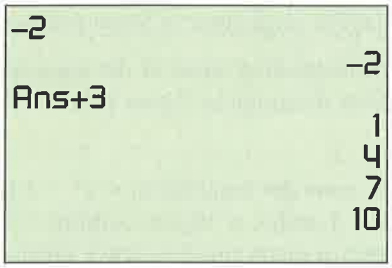
(a) (Explicit) $a_k = 3k - 5$ for $k = 1, 2, 3, \ldots$

(b) (Recursive) $a_1 = -2$ and $a_n = a_{n-1} + 3$ for $n = 2, 3, 4, \ldots$



```
0→K
                 0
K+1→K:3K-5
                -2
                 1
                 4
                 7
```

FIGURE 9.8 Typing these two commands (on the left of the viewing screen) will generate the terms of the explicitly defined sequence $a_k = 3k - 5$. (Example 7a)



```
-2
                -2
Ans+3
                 1
                 4
                 7
                10
```

FIGURE 9.9 Typing these two commands (on the left of the viewing screen) will generate the terms of the recursively defined sequence with $a_1 = -2$ and $a_n = a_{n-1} + 3$ (Example 7b).

SOLUTION

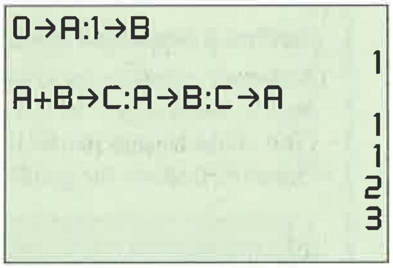
(a) On the home screen, type the two commands shown in Figure 9.8. The calculator will then generate the terms of the sequence as you push the ENTER key repeatedly.

(b) On the home screen, type the two commands shown in Figure 9.9. The first command gives the value of a_1. The calculator will generate the remaining terms of the sequence as you push the ENTER key repeatedly.

Notice that these two definitions generate the very same sequence!

Now try Exercises 1 and 5 on your calculator.

A recursive definition of a_n can be made in terms of any combination of preceding terms, just as long as those preceding terms have already been determined. A famous example is the **Fibonacci sequence**, named for Leonardo of Pisa (ca. 1170–1250), who wrote under the name Fibonacci. You can generate it with the two commands shown in Figure 9.10.



```
0→A:1→B
                 1
A+B→C:A→B:C→A
                 1
                 1
                 2
                 3
```

FIGURE 9.10 The two commands on the left will generate the Fibonacci sequence as the ENTER key is pressed repeatedly.

The Fibonacci sequence can be defined recursively using three statements.

The Fibonacci Sequence

The Fibonacci sequence can be defined recursively by

$$a_1 = 1$$
$$a_2 = 1$$
$$a_n = a_{n-2} + a_{n-1}$$

for all positive integers $n \geq 3$.

FIBONACCI NUMBERS

The numbers in the Fibonacci sequence have fascinated professional and amateur mathematicians alike since the thirteenth century. Not only is the sequence, like Pascal's triangle, a rich source of curious internal patterns, but the Fibonacci numbers seem to appear everywhere in nature. If you count the leaflets on a leaf, the leaves on a stem, the whorls on a pine cone, the rows on an ear of corn, the spirals in a sunflower, or the branches from a trunk of a tree, they tend to be Fibonacci numbers. (Check **phyllotaxy** in a biology book.)

Summation Notation

We want to look at the formulas for summing the terms of arithmetic and geometric sequences, but first we need a notation for writing the sum of an indefinite number of terms. The capital Greek letter sigma (Σ) provides our shorthand notation for a "summation."

Summations on a Calculator

If you think of summations as summing sequence values, it is not hard to translate sigma notation into calculator syntax. Here, in calculator syntax, are the first three summations in Exploration 1. (Don't try these on your calculator until you have first figured out the answers with pencil and paper.)

1. sum (seq (3K, K, 1, 5))
2. sum (seq (K^2, K, 5, 8))
3. sum (seq (cos (Nπ), N, 0, 12))

Definition Summation Notation

In **summation notation**, the sum of the terms of the sequence $\{a_1, a_2, \ldots, a_n\}$ is denoted

$$\sum_{k=1}^{n} a_k$$

which is read "the sum of a_k from $k = 1$ to n."

The variable k is called the **index of summation**.

EXPLORATION 1 Summing with Sigma

Sigma notation is actually even more versatile than the definition above suggests. See if you can determine the number represented by each of the following expressions.

1. $\sum_{k=1}^{5} 3k$ **2.** $\sum_{k=5}^{8} k^2$ **3.** $\sum_{n=0}^{12} \cos(n\pi)$ **4.** $\sum_{n=1}^{\infty} \sin(n\pi)$ **5.** $\sum_{k=1}^{\infty} \frac{3}{10^k}$

(If you're having trouble with number 5, here's a hint: Write the sum as a decimal!)

Although you probably computed them correctly, there is more going on in number 4 and number 5 in the above exploration than first meets the eye. We will have more to say about these "infinite" summations toward the end of this section.

Sums of Arithmetic and Geometric Sequences

One of the most famous legends in the lore of mathematics concerns the German mathematician Karl Friedrich Gauss (1777–1855), whose mathematical talent was apparent at a very early age. One version of the story has Gauss, at age ten, being in a class that was challenged by the teacher to add up all the numbers from 1 to 100. While his classmates were still writing down the problem, Gauss walked to the front of the room to present his slate to the teacher. The teacher, certain that Gauss could only be guessing, refused to look at his answer. Gauss simply placed it face down on the teacher's desk, declared "There it is," and returned to his seat. Later, after all the slates had been collected, the teacher looked at Gauss's work, which consisted of a single number: the correct answer. No other student (the legend goes) got it right.

The important feature of this legend for mathematicians is *how* the young Gauss got the answer so quickly. We'll let you reproduce his technique in Exploration 2.

EXPLORATION 2 Gauss's Insight

Your challenge is to find the sum of the natural numbers from 1 to 100 without a calculator.

1. On a wide piece of paper, write the sum "$1 + 2 + 3 + \cdots + 98 + 99 + 100$."
2. Underneath this sum, write the sum "$100 + 99 + 98 + \cdots + 3 + 2 + 1$."
3. Add the numbers two-by-two in *vertical* columns and notice that you get the same identical sum 100 times. What is it?
4. What is the sum of the 100 identical numbers referred to in part 3?
5. Explain why half the answer in part 4 is the answer to the challenge. Can you find it without a calculator?

If this story is true, then the youthful Gauss had discovered a fact that his elders knew about arithmetic sequences. If you write a finite arithmetic sequence forward on one line and backward on the line below it, then all the pairs stacked vertically sum to the same number. Multiplying this number by the number of terms n and dividing by 2 gives us a shortcut to the sum of the n terms. We state this result as a theorem.

Theorem Sum of a Finite Arithmetic Sequence

Let $\{a_1, a_2, a_3, \ldots, a_n\}$ be a finite arithmetic sequence with common difference d. Then the sum of the terms of the sequence is

$$\begin{aligned}\sum_{k=1}^{n} a_k &= a_1 + a_2 + \cdots + a_n\\ &= n\left(\frac{a_1 + a_n}{2}\right)\\ &= \frac{n}{2}(2a_1 + (n-1)d)\end{aligned}$$

Proof

We can construct the sequence forward by starting with a_1 and *adding* d each time, or we can construct the sequence backward by starting at a_n and *subtracting* d each time. We thus get two expressions for the sum we are looking for:

$$\sum_{k=1}^{n} a_k = a_1 + (a_1 + d) + (a_1 + 2d) + \cdots + (a_1 + (n-1)d)$$

$$\sum_{k=1}^{n} a_k = a_n + (a_n - d) + (a_n - 2d) + \cdots + (a_n - (n-1)d)$$

Summing vertically, we get

$$2\sum_{k=1}^{n} a_k = (a_1 + a_n) + (a_1 + a_n) + \cdots + (a_1 + a_n)$$

$$2\sum_{k=1}^{n} a_k = n(a_1 + a_n)$$

$$\sum_{k=1}^{n} a_k = n\left(\frac{a_1 + a_n}{2}\right)$$

If we substitute $a_1 + (n - 1)d$ for a_n, we get the alternate formula

$$\sum_{k=1}^{n} a_k = \frac{n}{2}(2a_1 + (n - 1)d).$$

EXAMPLE 8 Summing the terms of an arithmetic sequence

A corner section of a stadium has 8 seats along the front row. Each successive row has two more seats than the row preceding it. If the top row has 24 seats, how many seats are in the entire section?

SOLUTION The numbers of seats in the rows form an arithmetic sequence with

$$a_1 = 8, \quad a_n = 24, \quad \text{and} \quad d = 2.$$

Solving $a_n = a_1 + (n - 1)d$, we find that

$$\begin{aligned} 24 &= 8 + (n - 1)(2) \\ 16 &= (n - 1)(2) \\ 8 &= n - 1 \\ n &= 9 \end{aligned}$$

Applying the Sum of a Finite Arithmetic Sequence Theorem, the total number of seats in the section is $9(8 + 24)/2 = 144$.

We can support this answer numerically by computing the sum on a calculator:

$$\text{sum(seq}(8 + (\text{N} - 1)2, \text{N}, 1, 9) = 144.$$

Now try Exercise 33.

As you might expect, there is also a convenient formula for summing the terms of a finite geometric sequence.

Theorem Sum of a Finite Geometric Sequence

Let $\{a_1, a_2, a_3, \ldots, a_n\}$ be a finite geometric sequence with common ratio $r \neq 1$.

Then the sum of the terms of the sequence is

$$\begin{aligned}\sum_{k=1}^{n} a_k &= a_1 + a_2 + \cdots + a_n \\ &= \frac{a_1(1 - r^n)}{1 - r}\end{aligned}$$

Proof

Because the sequence is geometric, we have

$$\sum_{k=1}^{n} a_k = a_1 + a_1 \cdot r + a_1 \cdot r^2 + \cdots + a_1 \cdot r^{n-1}.$$

Therefore,

$$r \cdot \sum_{k=1}^{n} a_k = a_1 \cdot r + a_1 \cdot r^2 + \cdots + a_1 \cdot r^{n-1} + a_1 \cdot r^n.$$

If we now *subtract* the lower summation from the one above it, we have (after eliminating a lot of zeros):

$$\begin{aligned}\left(\sum_{k=1}^{n} a_k\right) - r \cdot \left(\sum_{k=1}^{n} a_k\right) &= a_1 - a_1 \cdot r^n \\ \left(\sum_{k=1}^{n} a_k\right)(1 - r) &= a_1(1 - r^n) \\ \sum_{k=1}^{n} a_k &= \frac{a_1(1 - r^n)}{1 - r}\end{aligned}$$

EXAMPLE 9 Summing the terms of a geometric sequence

Find the sum of the geometric sequence $4, -4/3, 4/9, -4/27, \ldots, 4(-1/3)^{10}$.

SOLUTION We can see that $a_1 = 4$ and $r = -1/3$. The nth term is $4(-1/3)^{10}$, which means that $n = 11$. (Remember that the exponent on the nth term is $n - 1$, not n.) Applying the Sum of a Finite Geometric Sequence Theorem, we find that

$$\sum_{n=1}^{11} 4\left(-\frac{1}{3}\right)^{n-1} = \frac{4(1 - (-1/3)^{11})}{1 - (-1/3)} \approx 3.000016935.$$

We can support this answer by having the calculator do the actual summing:

sum(seq(4(−1/3)^(N − 1), N, 1, 11) = 3.000016935.

Now try Exercise 39.

As one practical application of the Sum of a Finite Geometric Sequence Theorem, we will tie up a loose end from Section 3.6, wherein you learned that the future value *FV* of an ordinary annuity consisting of *n* equal periodic sums." For example, we can add any finite number of 2's together and get a real number, but if we add an *infinite* number of 2's together we do not get a real number at all. Sums do not behave that way.

What makes series so interesting is that sometimes (as in Example 9) the sequence of **partial sums**, all of which are true sums, approaches a finite limit S:

$$\lim_{n\to\infty}\sum_{k=1}^{n} a_k = \lim_{n\to\infty}(a_1 + a_2 + \cdots + a_n) = S.$$

In this case we say that the series **converges** to S, and it makes sense to define S as the **sum of the infinite series**. In sigma notation,

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty}\sum_{k=1}^{n} a_k = S.$$

If the limit of partial sums does not exist, then the series **diverges** and has no sum.

EXAMPLE 10 Looking at limits of partial sums

For each of the following series, find the first five terms in the sequence of partial sums. Which of the series appear to converge?

(a) $0.1 + 0.01 + 0.001 + 0.0001 + \cdots$

(b) $10 + 20 + 30 + 40 + \cdots$

(c) $1 - 1 + 1 - 1 + \cdots$

SOLUTION

(a) The first five partial sums are $\{0.1, 0.11, 0.111, 0.1111, 0.11111\}$. These appear to be approaching a limit of $0.\overline{1} = 1/9$, which would suggest that the series converges to a sum of $1/9$.

(b) The first five partial sums are $\{10, 30, 60, 100, 150\}$. These numbers increase without bound and do not approach a limit. The series diverges and has no sum.

(c) The first five partial sums are $\{1, 0, 1, 0, 1\}$. These numbers oscillate and do not approach a limit. The series diverges and has no sum.

Now try Exercise 49.

You might have been tempted to "pair off" the terms in Example 10c to get an infinite summation of 0's (and hence a sum of 0), but you would be applying a rule (namely the *associative property of addition*) that works on *finite* sums but not, in general, on infinite series. The sequence of partial sums does not have a limit, so any manipulation of the series in Example 10c that appears to result in a sum is actually meaningless.

payments of R dollars at an interest rate i per compounding period (payment interval) is

$$FV = R\frac{(1+i)^n - 1}{i}.$$

We can now consider the mathematics behind this formula. The n payments remain in the account for different lengths of time and so earn different amounts of interest. The total value of the annuity after n payment periods (see Example 8 in Section 3.6) is

$$FV = R + R(1+i) + R(1+i)^2 + \cdots + R(1+i)^{n-1}.$$

The terms of this sum form a geometric sequence with first term R and common ratio $(1 + i)$. Applying the Sum of a Finite Geometric Sequence Theorem, the sum of the n terms is

$$\begin{aligned} FV &= \frac{R(1-(1+i)^n)}{1-(1+i)} \\ &= R\frac{1-(1+i)^n}{-i} \\ &= R\frac{(1+i)^n - 1}{i} \end{aligned}$$

Infinite Series

If you change the "11" in the calculator sum in Example 9 to higher and higher numbers, you will find that the sum approaches a value of 3. This is no coincidence. In the language of limits,

$$\begin{aligned} \lim_{n\to\infty}\sum_{k=1}^{n} 4\left(-\frac{1}{3}\right)^{k-1} &= \lim_{n\to\infty}\frac{4(1-(-1/3)^n)}{1-(-1/3)} \\ &= \frac{4(1-0)}{4/3} \qquad \text{Since } \lim_{n\to\infty}(-1/3)^n = 0. \\ &= 3 \end{aligned}$$

This gives us the opportunity to extend the usual meaning of the word "sum," which always applies to a *finite* number of terms being added together. By using limits, we can make sense of expressions in which an *infinite* number of terms are added together. Such expressions are called **infinite series**.

Definition Infinite Series

An **infinite series** is an expression of the form

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + \cdots + a_n + \cdots.$$

The first thing to understand about an infinite series is that it is not a true sum. There are properties of real number addition that allow us to extend the definition of $a + b$ to sums like $a + b + c + d + e + f$, but not to "infinite

Convergence of Geometric Series

Determining the convergence or divergence of infinite series is an important part of a calculus course, in which series are used to represent functions. Most of the convergence tests are well beyond the scope of this course, but we are in a position to settle the issue completely for geometric series.

Theorem Sum of an Infinite Geometric Series

The geometric series $\sum_{k=1}^{\infty} a \cdot r^{k-1}$ converges if and only if $|r| < 1$. If it does converge, the sum is $a/(1 - r)$.

Proof

If $r = 1$, the series is $a + a + a + \cdots$, which is unbounded and hence diverges.

If $r = -1$, the series is $a - a + a - a + \cdots$, which diverges. (See Example 10c.)

If $r \neq 1$, then by the Sum of a Finite Geometric Sequence Theorem, the nth partial sum of the series is $\sum_{k=1}^{n} a \cdot r^{k-1} = a(1 - r^n)/(1 - r)$. The limit of the partial sums is $\lim_{n\to\infty}[a(1 - r^n)/(1 - r)]$, which converges if and only if $\lim_{n\to\infty} r^n$ is a finite number. But $\lim_{n\to\infty} r^n$ is 0 when $|r| < 1$ and unbounded when $|r| > 1$. Therefore, the sequence of partial sums converges if and only if $|r| < 1$, in which case the sum of the series is

$$\lim_{n\to\infty}[a(1 - r^n)/(1 - r)] = a(1 - 0)/(1 - r) = a/(1 - r).$$

EXAMPLE 11 Summing infinite geometric series

Determine whether the series converges. If it converges, give the sum.

(a) $\sum_{k=1}^{\infty} 3(0.75)^{k-1}$ **(b)** $\sum_{n=0}^{\infty} \left(-\frac{4}{5}\right)^n$

(c) $\sum_{n=1}^{\infty} \left(\frac{\pi}{2}\right)^n$ **(d)** $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$

SOLUTION

(a) Since $|r| = |0.75| < 1$, the series converges. The first term is $3(0.75)^0 = 3$, so the sum is $a/(1 - r) = 3/(1 - 0.75) = 12$.

(b) Since $|r| = |-4/5| < 1$, the series converges. The first term is $(-4/5)^0 = 1$, so the sum is $a/(1 - r) = 1/(1 - (-4/5)) = 5/9$.

(c) Since $|r| = |\pi/2| > 1$, the series diverges.

(d) Since $|r| = |1/2| < 1$, the series converges. The first term is 1, and so the sum is $a/(1 - r) = 1/(1 - 1/2) = 2$.

Now try Exercise 51.

EXAMPLE 12 Converting a repeating decimal to fraction form

Express $0.\overline{234} = 0.234234234\ldots$ in fraction form.

SOLUTION We can write this number as a sum: $0.234 + 0.000234 + 0.000000234 + \cdots$.

This is a convergent infinite geometric series in which $a = 0.234$ and $r = 0.001$. The sum is

$$\frac{a}{1-r} = \frac{0.234}{1-0.001} = \frac{0.234}{0.999} = \frac{234}{999} = \frac{26}{111}.$$

Now try Exercise 57.

QUICK REVIEW 9.4

(For help, see Section P.1.)

In Exercises 1 and 2, evaluate each expression when $a = 3$, $d = 4$, and $n = 5$.

1. $a + (n-1)d$

2. $\frac{n}{2}(2a + (n-1)d)$

In Exercises 3 and 4, evaluate each expression when $a = 5$, $r = 4$, and $n = 3$.

3. $a \cdot r^{n-1}$

4. $\frac{a(1-r^n)}{1-r}$

In Exercises 5–10, find a_{10}.

5. $a_k = \frac{k}{k+1}$

6. $a_k = 5 + (k-1)3$

7. $a_k = 5 \cdot 2^{k-1}$

8. $a_k = \frac{4}{3}\left(\frac{1}{2}\right)^{k-1}$

9. $a_k = 32 - a_{k-1}$ and $a_9 = 17$

10. $a_k = \frac{k^2}{2^k}$

$2+(n-1)(3-n)$

SECTION 9.4 EXERCISES

In Exercises 1–4, find the first 6 terms and the 100th term of the explicitly-defined sequence.

1. $u_n = \frac{n+1}{n}$

2. $v_n = \frac{4}{n+2}$

3. $c_n = n^3 - n$

4. $d_n = n^2 - 5n$

In Exercises 5–10, find the first 4 terms and the eighth term of the recursively-defined sequence.

5. $a_1 = 8$ and $a_n = a_{n-1} - 4$, for $n \geq 2$

6. $u_1 = -3$ and $u_{k+1} = u_k + 10$, for $k \geq 1$

7. $b_1 = 2$ and $b_{k+1} = 3b_k$, for $k \geq 1$

8. $v_1 = 0.75$ and $v_n = (-2)v_{n-1}$, for $n \geq 2$

9. $c_1 = 2$, $c_2 = -1$, and $c_{k+2} = c_k + c_{k+1}$, for $k \geq 1$

10. $c_1 = -2$, $c_2 = 3$, and $c_k = c_{k-2} + c_{k-1}$, for $k \geq 3$

In Exercises 11–14, the sequences are arithmetic. Find

(a) the common difference,

(b) the tenth term,

(c) a recursive rule for the *n*th term, and

(d) an explicit rule for the *n*th term.

11. 6, 10, 14, 18, . . .

12. −4, 1, 6, 11, . . .

13. −5, −2, 1, 4, . . .

14. −7, 4, 15, 26, . . .

In Exercises 15–18, the sequences are geometric. Find

(a) the common ratio,

(b) the eighth term,

(c) a recursive rule for the *n*th term, and

(d) an explicit rule for the *n*th term.

15. 2, 6, 18, 54, . . .

16. 3, 6, 12, 24, . . .

17. 1, −2, 4, −8, 16, . . .

18. −2, 2, −2, 2, . . .

19. The fourth and seventh terms of an arithmetic sequence are −8 and 4, respectively. Find the first term and a recursive rule for the *n*th term.

20. The fifth and ninth terms of an arithmetic sequence are −5 and −17, respectively. Find the first term and a recursive rule for the *n*th term.

21. The second and eighth terms of a geometric sequence are 3 and 192, respectively. Find the first term, common ratio, and an explicit rule for the *n*th term.

22. The third and sixth terms of a geometric sequence are -75 and -9375, respectively. Find the first term, common ratio, and an explicit rule for the nth term.

In Exercises 23–26, graph the sequence.

23. $a_n = 2 - \dfrac{1}{n}$ **24.** $b_n = \sqrt{n} - 3$

25. $c_n = n^2 - 5$ **26.** $d_n = 3 + 2n$

In Exercises 27–32, write each sum using summation notation, assuming the suggested pattern continues.

27. $-7 - 1 + 5 + 11 + \cdots + 53$

28. $2 + 5 + 8 + 11 + \cdots + 29$

29. $1 + 4 + 9 + \cdots + (n + 1)^2$

30. $1 + 8 + 27 + \cdots + (n + 1)^3$

31. $6 - 12 + 24 - 48 + \cdots$

32. $5 - 15 + 45 - 135 + \cdots$

In Exercises 33–38, find the sum of the arithmetic sequence.

33. $-7, -3, 1, 5, 9, 13$

34. $-8, -1, 6, 13, 20, 27$

35. $1, 2, 3, 4, \ldots, 80$

36. $2, 4, 6, 8, \ldots, 70$

37. $117, 110, 103, \ldots, 33$

38. $111, 108, 105, \ldots, 27$

In Exercises 39–42, find the sum of the geometric sequence.

39. $3, 6, 12, \ldots, 12{,}288$

40. $5, 15, 45, \ldots, 98{,}415$

41. $42, 7, \dfrac{7}{6}, \ldots, 42\left(\dfrac{1}{6}\right)^8$

42. $42, -7, \dfrac{7}{6}, \ldots, 42\left(-\dfrac{1}{6}\right)^9$

In Exercises 43–48, find the sum of the first n terms of the sequence. The sequence is either arithmetic or geometric.

43. $2, 5, 8, \ldots; n = 10$ **44.** $14, 8, 2, \ldots; n = 9$

45. $4, -2, 1, -\dfrac{1}{2}, \ldots; n = 12$

46. $6, -3, \dfrac{3}{2}, -\dfrac{3}{4}, \ldots; n = 11$

47. $-1, 11, -121, \ldots; n = 9$

48. $-2, 24, -288, \ldots; n = 8$

49. Find the first six partial sums of the following infinite series. If the sums have a finite limit, write "convergent." If not, write "divergent."

(a) $0.3 + 0.03 + 0.003 + 0.0003 + \cdots$

(b) $1 - 2 + 3 - 4 + 5 - 6 + \cdots$

50. Find the first six partial sums of the following infinite series. If the sums have a finite limit, write "convergent." If not, write "divergent."

(a) $-2 + 2 - 2 + 2 - 2 + \cdots$

(b) $1 - 0.7 - 0.07 - 0.007 - 0.0007 - \cdots$

In Exercises 51–56, determine whether the infinite geometric series converges. If it does, find its sum.

51. $6 + 3 + \dfrac{3}{2} + \dfrac{3}{4} + \cdots$ **52.** $4 + \dfrac{4}{3} + \dfrac{4}{9} + \dfrac{4}{27} + \cdots$

53. $\dfrac{1}{64} + \dfrac{1}{32} + \dfrac{1}{16} + \dfrac{1}{8} + \cdots$

54. $\dfrac{1}{48} + \dfrac{1}{16} + \dfrac{3}{16} + \dfrac{9}{16} + \cdots$

55. $\displaystyle\sum_{j=1}^{\infty} 3\left(\frac{1}{4}\right)^j$ **56.** $\displaystyle\sum_{n=1}^{\infty} 5\left(\frac{2}{3}\right)^n$

In Exercises 57–60, express the rational number as a fraction of integers.

57. $7.14141414\ldots$ **58.** $5.93939393\ldots$

59. $-17.268268268\ldots$ **60.** $-12.876876876\ldots$

61. Rain Forest Growth The bungy-bungy tree in the Amazon rain forest grows an average 2.3 cm per week. Write a sequence that represents the weekly height of a bungy-bungy over the course of 1 year if it is 7 meters tall today. Display the first four terms and the last two terms.

62. Half-Life (See Section 3.2) Thorium-232 has a half-life of 14 billion years. Make a table showing the half-life decay of a sample of Thorium-232 from 16 grams to 1 gram; list the time (in years, starting with $t = 0$) in the first column and the mass (in grams) in the second column. Which type of sequence is each column of the table?

63. Savings Account The table below shows the December balance in a fixed-rate compound savings account each year from 1996 to 2000.

Year	1996	1997	1998	1999	2000
Balance	\$20,000	\$22,000	\$24,200	\$26,620	\$29,282

(a) The balances form a geometric sequence. What is r?

(b) Write a formula for the balance in the account n years after December 1996.

(c) Find the sum of the December balances from 1996 to 2006, inclusive.

64. Savings Account The table below shows the December balance in a simple interest savings account each year from 1996 to 2000.

Year	1996	1997	1998	1999	2000
Balance	\$18,000	\$20,016	\$22,032	\$24,048	\$26,064

(a) The balances form an arithmetic sequence. What is d?

(b) Write a formula for the balance in the account n years after December 1996.

(c) Find the sum of the December balances from 1996 to 2006, inclusive.

65. Annuity Mr. O'Hara deposits \$120 at the end of each month into an account that pays 7% interest compounded monthly. After 10 years, the balance in the account, in dollars, is

$$120\left(1+\frac{0.07}{12}\right)^0+120\left(1+\frac{0.07}{12}\right)^1+\cdots+120\left(1+\frac{0.07}{12}\right)^{119}.$$

(a) This is a geometric series. What is the first term? What is r?

(b) Use the formula for the sum of a finite geometric sequence to find the balance.

66. Annuity Ms. Argentieri deposits \$100 at the end of each month into an account that pays 8% interest compounded monthly. After 10 years, the balance in the account, in dollars, is

$$100\left(1+\frac{0.08}{12}\right)^0+100\left(1+\frac{0.08}{12}\right)^1+\cdots+100\left(1+\frac{0.08}{12}\right)^{119}.$$

(a) This is a geometric series. What is the first term? What is r?

(b) Use the formula for the sum of a finite geometric sequence to find the balance.

67. Arena Seating The first row of seating in section J of the Athena Arena has 7 seats. In all, there are 25 rows of seats in section J, each row containing two more seats than the row preceding it. How many seats are in section J?

68. Patio Construction Pat designs a patio with a trapezoid-shaped deck consisting of 16 rows of congruent slate tiles. The numbers of tiles in the rows form an arithmetic sequence. The first row contains 15 tiles and the last row contains 30 tiles. How many tiles are used in the deck?

69. Group Activity Follow the Bouncing Ball When "superballs" sprang upon the scene in the 1960s, kids across the United States were amazed that these hard rubber balls could bounce to 90% of the height from which they were dropped. If a superball is dropped from a height of 2 m, how far does it travel by the time it hits the ground for the tenth time? [*Hint:* The ball goes down to the first bounce, then up *and* down thereafter.]

70. Group Activity End Behavior Does the graph of

$$f(x)=2\left(\frac{1-1.05^x}{1-1.05}\right)$$

have a horizontal asymptote to the right? How does this relate to the convergence or divergence of the series

$$2+2.1+2.205+2.31525+\cdots?$$

Standardized Test Questions

71. True or False If the first two terms of a geometric sequence are negative, then so is the third. Justify your answer.

72. True or False If the first two terms of an arithmetic sequence are positive, then so is the third. Justify your answer.

You may use a graphing calculator when solving Exercises 73–76.

73. The first two terms of an arithmetic sequence are 2 and 8. The fourth term is

(a) 20.

(b) 26.

(c) 64.

(d) 128.

(e) 256.

74. The sum of an infinite geometric series with first term 3 and second term 0.75 is

(a) 3.75.

(b) 2.4.

(c) 4.

(d) 5.

(e) 12.

75. $\sum_{n=1}^{5} n^2 =$

(a) 25

(b) 26

(c) 30

(d) 55

(e) 225

76. $\sum_{n=0}^{\infty} 4\left(-\frac{5}{3}\right)^n =$

(a) -6

(b) $-\frac{5}{2}$

(c) $\frac{3}{2}$

(d) 10

(e) divergent

Explorations

77. Rabbit Populations Assume that 2 months after birth, each male-female pair of rabbits begins producing one new male-female pair of rabbits each month. Further assume that the rabbit colony begins with one newborn male-female pair of rabbits and no rabbits die for 12 months. Let a_n represent the number of *pairs* of rabbits in the colony after $n - 1$ months.

(a) Writing to Learn Explain why $a_1 = 1$, $a_2 = 1$, and $a_3 = 2$.

(b) Find $a_4, a_5, a_6, \ldots, a_{13}$.

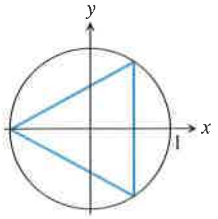
(c) Writing to Learn Explain why the sequence $\{a_n\}$, $1 \leq n \leq 13$, is a model for the size of the rabbit colony for a 1-year period.

78. Fibonacci Sequence Compute the first seven terms of the sequence whose nth term is

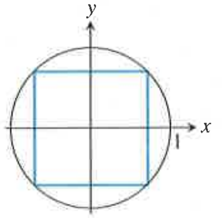
$$a_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

How do these seven terms compare with the first seven terms of the Fibonacci sequence?

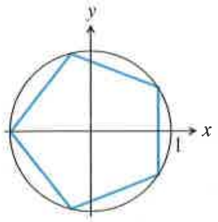
79. Connecting Geometry and Sequences In the following sequence of diagrams, regular polygons are inscribed in unit circles with at least one side of each polygon perpendicular to the positive x-axis.



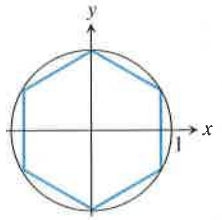
(a)



(b)



(c)



(d)

(a) Prove that the perimeter of each polygon in the sequence is given by $a_n = 2n \sin(\pi/n)$, where n is the number of sides in the polygon.

(b) Investigate the value of a_n for $n = 10$, 100, 1000, and 10,000. What conclusion can you draw?

80. Recursive Sequence The population of Centerville was 525,000 in 1992 and is growing annually at the rate of 1.75%. Write a recursive sequence $\{P_n\}$ for the population. State the first term P_1 for your sequence.

81. Writing to Learn If $\{a_n\}$ is a geometric sequence with all positive terms, explain why $\{\log a_n\}$ must be arithmetic.

82. Writing to Learn If $\{b_n\}$ is an arithmetic sequence, explain why $\{10^{b_n}\}$ must be geometric.

83. Population Density The *National Geographic Picture Atlas of Our Fifty States* (2001) groups the states into 10 regions. The two largest groupings are the Heartland (Table 9.1) and the Southeast (Table 9.2). Population and area data for the two regions are given in the tables. The populations are official 2000 U.S. Census figures.

(a) What is the total population of each region?

(b) What is the total area of each region?

(c) What is the population density (in persons per square mile) of each region?

(d) Writing to Learn For the two regions, compute the population density of each state. What is the average of the seven state population densities for each region? Explain why these answers differ from those found in part (c).

TABLE 9.1 THE HEARTLAND

State	Population	Area (mi^2)
Iowa	2,926,324	56,275
Kansas	2,688,418	82,277
Minnesota	4,919,479	84,402
Missouri	5,595,211	69,697
Nebraska	1,711,283	77,355
North Dakota	642,200	70,703
South Dakota	754,844	77,116

TABLE 9.2 THE SOUTHEAST

State	Population	Area (mi^2)
Alabama	4,447,100	51,705
Arkansas	2,673,400	53,187
Florida	15,982,378	58,644
Georgia	8,186,453	58,910
Louisiana	4,468,976	47,751
Mississippi	2,844,658	47,689
S. Carolina	4,012,012	31,113

84. Finding a Pattern Write the finite series $-1 + 2 + 7 + 14 + 23 + \cdots + 62$ in summation notation.

Extending the Ideas

85. A Sequence of Matrices Write out the first seven terms of the "geometric sequence" with the first term the matrix $\begin{bmatrix} 1 & 1 \end{bmatrix}$ and the common ratio the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. How is this sequence of matrices related to the Fibonacci sequence?

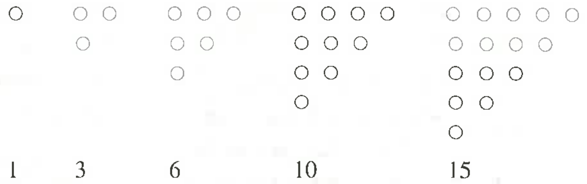
86. Prove that the sum of the first n odd positive integers is n^2.

87. Fibonacci Sequence and Series Complete the following table, where F_n is the nth term of the Fibonacci sequence and S_n is the nth partial sum of the Fibonacci series. Make a conjecture based on the numerical evidence in the table.

$$S_n = \sum_{k=1}^{n} F_k$$

n	F_n	S_n	$F_{n+2} - 1$
1	1		
2	1		
3	2		
4			
5			
6			
7			
8			
9			

88. Triangular Numbers Revisited Exercise 41 in Section 9.2 introduced **triangular numbers** as numbers that count objects arranged in triangular arrays:



In that exercise, you gave a geometric argument that the nth triangular number was $n(n + 1)/2$. Prove that formula algebraically using the Sum of a Finite Arithmetic Sequence Theorem.

89. Square Numbers and Triangular Numbers Prove that the sum of two consecutive triangular numbers is a square number; that is, prove

$$T_{n-1} + T_n = n^2$$

for all positive integers $n \geq 2$. Use both a geometric and an algebraic approach.

90. Harmonic Series Graph the sequence of partial sums of the *harmonic series:*

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots.$$

Overlay on it the graph of $f(x) = \ln x$. The resulting picture should support the claim that

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} \geq \ln n,$$

for all positive integers n. Make a table of values to further support this claim. Explain why the claim implies that the harmonic series must diverge.

9.5 MATHEMATICAL INDUCTION

What you'll learn about

- The Tower of Hanoi Problem
- Principle of Mathematical Induction
- Induction and Deduction

. . . and why

The principle of mathematical induction is a valuable technique for proving combinatorial formulas.

The Tower of Hanoi Problem

You might be familiar with a game that is played with a stack of round washers of different diameters and a stand with three vertical pegs (Figure 9.11). The game is not difficult to win once you get the hang of it, but it takes a while to move all the washers even when you know what you are doing. A mathematician, presented with this game, wants to figure out the *minimum* number of moves required to win the game—not because of impatience, but because it is an interesting mathematical problem.

In case mathematics is not sufficient motivation to look into the problem, there is a legend attached to the game that provides a sense of urgency. The legend has it that a game of this sort with 64 golden washers was created at the beginning of time. A special order of far eastern monks has been moving the washers at one move per second ever since, always using the minimum number of moves required to win the game. When the final washer is moved, that will be the end of time. The Tower of Hanoi Problem is simply to figure out how much time we have left.

We will solve the problem by proving a general theorem that gives the minimum number of moves for any number of washers. The technique of proof we use is called the principle of mathematical induction, the topic of this section.

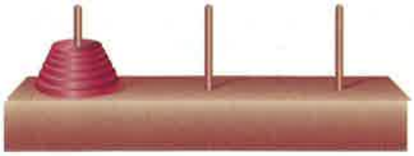


FIGURE 9.11 The Tower of Hanoi Game. The object is to move the entire stack of washers to the rightmost peg, one washer at a time, never placing a larger washer on top of a smaller washer.

TOWER OF HANOI HISTORY

The legend notwithstanding, the Tower of Hanoi dates back to 1883, when Édouard Lucas marketed the game as "La Tour de Hanoï," brought back from the Orient by "Professor N. Claus de Siam"—an anagram of "Professor Lucas d'Amiens." The legend appeared shortly thereafter. The game has been a favorite among computer programmers, so a web search on "Tower of Hanoi" will bring up multiple sites that allow you to play it on your home computer.

Theorem The Tower of Hanoi Solution

The minimum number of moves required to move a stack of n washers in a Tower of Hanoi game is $2^n - 1$.

Proof

(The Anchor) First, we note that the assertion is true when $n = 1$. We can certainly move the one washer to the right peg in (minimally) one move, and $2^1 - 1 = 1$.

(The Inductive Hypothesis) Now let us assume that the assertion holds for $n = k$; that is, the minimum number of moves required to move k washers is $2^k - 1$. (So far the only k we are sure of is 1, but keep reading.)

(The Inductive Step) We next consider the case when $n = k + 1$ washers. To get at the bottom washer, we must first move the entire stack of k washers sitting on top of it. *By the assumption we just made*, this will take a minimum of $2^k - 1$ moves. We can then move the bottom washer to the free peg (1 move). Finally, we must move the stack of k washers back onto the bottom washer—again, *by our assumption*, a minimum of $2^k - 1$ moves. Altogether, moving $k + 1$ washers requires

$$(2^k - 1) + 1 + (2^k - 1) = 2 \cdot 2^k - 1 = 2^{k+1} - 1$$

moves. Since that agrees with the formula in the statement of the proof, we have shown the assertion to be true for $n = k + 1$ washers—under the assumption that it is true for $n = k$.

Remarkably, we are done. Recall that we *did* prove the theorem to be true for $n = 1$. Therefore, by the Inductive Step, it must also be true for $n = 2$. By the Inductive Step again, it must be true for $n = 3$. And so on, for all positive integers n.

If we apply the Tower of Hanoi Solution to the legendary Tower of Hanoi Problem, the monks will need $2^{64} - 1$ seconds to move the 64 golden washers. The largest current conjecture for the age of the universe is something on the order of 20 billion years. If you convert $2^{64} - 1$ seconds to years, you will find that the end of time (at least according to this particular legend) is not exactly imminent. In fact, you might be surprised at how much time is left!

EXPLORATION 1 Winning the Game

One thing that the Tower of Hanoi Solution does not settle is how to get the stack to finish on the rightmost peg rather than the middle peg. Predictably, it depends on where you move the first washer, but it also depends on the height of the stack. Using a website game, or coins of different sizes, or even the real game if you have one, play the game with 1 washer, then 2, then 3, then 4, and so on, keeping track of what your first move must be in order to have the stack wind up on the rightmost peg in $2^n - 1$ moves. What is the general rule for a stack of n washers?

Principle of Mathematical Induction

The proof of the Tower of Hanoi Solution used a general technique known as the principle of mathematical induction. It is a powerful tool for proving all kinds of theorems about positive integers. We *anchor* the proof by establishing the truth of the theorem for 1, then we show the *inductive hypothesis* that "true for k" implies "true for $k + 1$."

Principle of Mathematical Induction

Let P_n be a statement about the integer n. Then P_n is true for all positive integers n provided the following conditions are satisfied:

1. (the anchor) P_1 is true;

2. (the inductive step) if P_k is true, then P_{k+1} is true.

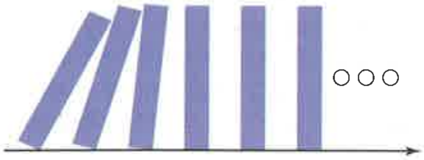


FIGURE 9.12 The Principle of Mathematical Induction visualized by dominoes. The toppling of domino #1 guarantees the toppling of domino n for all positive integers n.

A good way to visualize how the principle works is to imagine an infinite sequence of dominoes stacked upright, each one close enough to its neighbor so that any kth domino, if it falls, will knock over the $(k + 1)$st domino (Figure 9.12). Given that fact (the inductive step), the toppling of domino 1 guarantees the toppling of the entire infinite sequence of dominoes.

Let us use the principle to prove a fact that we already know.

EXAMPLE 1 Using mathematical induction

Prove that $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ is true for all positive integers n.

SOLUTION Call the statement P_n. We could verify P_n by using the formula for the sum of an arithmetic sequence, but here is how we prove it by mathematical induction.

(The Anchor) For $n = 1$, the equation reduces to P_1: $1 = 1^2$, which is true.

(The Inductive Hypothesis) Assume that the equation is true for $n = k$. That is, assume

$$P_k: 1 + 3 + \cdots + (2k - 1) = k^2 \text{ is true.}$$

(The Inductive Step) The next term on the left-hand side would be $2(k+1)-1$. We add this to both sides of P_k and get

$$\begin{aligned} 1 + 3 + \cdots + (2k-1) + (2(k+1)-1) &= k^2 + (2(k+1)-1) \\ &= k^2 + 2k + 1 \\ &= (k+1)^2 \end{aligned}$$

This is exactly the statement P_{k+1}, so the equation is true for $n = k+1$. Therefore, P_n is true for all positive integers, by mathematical induction.

Now try Exercise 1.

Notice that we did *not* plug in $k+1$ on both sides of the equation P_n in order to verify the inductive step; if we had done that, there would have been nothing to verify. If you find yourself verifying the inductive step without using the inductive hypothesis, you can assume that you have gone astray.

EXAMPLE 2 Using mathematical induction

Prove that $1^2 + 2^2 + 3^2 + \cdots + n^2 = [n(n+1)(2n+1)]/6$ is true for all positive integers n.

SOLUTION Let P_n be the statement $1^2 + 2^2 + 3^2 + \cdots + n^2 = [n(n+1)(2n+1)]/6$.

(The Anchor) P_1 is true because $1^2 = [1(2)(3)]/6$.

(The Inductive Hypothesis) Assume that P_k is true, so that

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}.$$

(The Inductive Step) The next term on the left-hand side would be $(k+1)^2$. We add this to both sides of P_k and get

$$\begin{aligned} 1^2 + 2^2 + \cdots + k^2 + (k+1)^2 &= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\ &= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\ &= \frac{(k+1)(2k^2 + k + 6k + 6)}{6} \\ &= \frac{(k+1)(k+2)(2k+3)}{6} \\ &= \frac{(k+1)((k+1)+1)(2(k+1)+1)}{6} \end{aligned}$$

This is exactly the statement P_{k+1}, so the equation is true for $n = k+1$. Therefore, P_n is true for all positive integers, by mathematical induction.

Now try Exercise 13.

Applications of mathematical induction can be quite different from the first two examples. Here is one involving divisibility.

EXAMPLE 3 Proving divisibility

Prove that $4^n - 1$ is evenly divisible by 3 for all positive integers n.

SOLUTION Let P_n be the statement that $4^n - 1$ is evenly divisible by 3 for all positive integers n.

(The Anchor) P_1 is true because $4^1 - 1 = 3$ is divisible by 3.

(The Inductive Hypothesis) Assume that P_k is true, so that $4^k - 1$ is divisible by 3.

(The Inductive Step) We need to prove that $4^{k+1} - 1$ is divisible by 3.

Using a little algebra, we see that $4^{k+1} - 1 = 4 \cdot 4^k - 1 = 4(4^k - 1) + 3$.

By the inductive hypothesis, $4^k - 1$ is divisible by 3. Of course, so is 3. Therefore, $4(4^k - 1) + 3$ is a sum of multiples of 3, and hence divisible by 3. This is exactly the statement P_{k+1}, so P_{k+1} is true. Therefore, P_n is true for all positive integers, by mathematical induction.

Now try Exercise 19.

$k^2+k+2(k+1)=(k+1)^2+(k+1)$

Induction and Deduction

The words *induction* and *deduction* are usually used to contrast two patterns of logical thought. We reason by **induction** when we use evidence derived from particular examples to draw conclusions about general principles. We reason by **deduction** when we reason from general principles to draw conclusions about specific cases.

When mathematicians prove theorems, they use deduction. In fact, even a "proof by mathematical induction" is a deductive proof, since it consists of applying the general principle to a particular formula. We have been careful to use the term **mathematical induction** in this section to distinguish it from inductive reasoning, which is often good for inspiring conjectures—but not for proving general principles.

Exploration 2 illustrates why mathematicians do not rely on inductive reasoning.

$2^x+2^x \Rightarrow 2^{(x+1)}$

EXPLORATION 2 Is $n^2 + n + 41$ Prime for all n?

1. Plug in the numbers from 1 to 10. Are the results all prime?
2. Repeat for the numbers from 11 to 20.
3. Repeat for the numbers from 21 to 30. (Ready to make your conjecture?)
4. What is the smallest value of n for which $n^2 + n + 41$ is not prime?

THE FOUR-COLOR MAP THEOREM

In 1852, Francis Guthrie conjectured that any map on a flat surface could be colored in at most four colors so that no two bordering regions would share the same color. Mathematicians tried unsuccessfully for almost 150 years to prove (or disprove) the conjecture, until Kenneth Appel and Wolfgang Haken finally proved it in 1976.

There is one situation in which (nonmathematical) induction can constitute a proof. In **enumerative induction**, one reasons from specific cases to the general principle by considering *all possible cases*. This is simple enough when proving a theorem like "All one-digit prime numbers are factors of 210," but it can involve some very elegant mathematics when the number of cases is seemingly infinite. Such was the case in the proof of the Four-Color Map Theorem, in which all possible cases were settled with the help of a clever computer program.

QUICK REVIEW 9.5

(Prerequisite skill Sections A.2, 1.2)

In Exercises 1–3, expand the product.

1. $n(n + 5)$

2. $(n + 2)(n - 3)$

3. $k(k + 1)(k + 2)$

In Exercises 4–6, factor the polynomial.

4. $n^2 + 2n - 3$

5. $k^3 + 3k^2 + 3k + 1$

6. $n^3 - 3n^2 + 3n - 1$

In Exercises 7–10, evaluate the function at the given domain values or variable expressions.

7. $f(x) = x + 4;\ f(1), f(t), f(t + 1)$

8. $f(n) = \dfrac{n}{n + 1};\ f(1), f(k), f(k + 1)$

9. $P(n) = \dfrac{2n}{3n + 1};\ P(1), P(k), P(k + 1)$

10. $P(n) = 2n^2 - n - 3;\ P(1), P(k), P(k + 1)$

SECTION 9.5 EXERCISES

In Exercises 1–4, use mathematical induction to prove that the statement holds for all positive integers.

1. $2 + 4 + 6 + \cdots + 2n = n^2 + n$

2. $8 + 10 + 12 + \cdots + (2n + 6) = n^2 + 7n$

3. $6 + 10 + 14 + \cdots + (4n + 2) = n(2n + 4)$

4. $14 + 18 + 22 + \cdots + (4n + 10) = 2n(n + 6)$

In Exercises 5–8, state an explicit rule for the nth term of the recursively defined sequence. Then use mathematical induction to prove the rule.

5. $a_n = a_{n-1} + 5,\ a_1 = 3$

6. $a_n = a_{n-1} + 2,\ a_1 = 7$

7. $a_n = 3a_{n-1},\ a_1 = 2$

8. $a_n = 5a_{n-1},\ a_1 = 3$

In Exercises 9–12, write the statements P_1, P_k, and P_{k+1}. (Do not write a proof.)

9. P_n: $1 + 2 + \cdots + n = \dfrac{n(n + 1)}{2}$

10. P_n: $1^2 + 3^2 + 5^2 + \cdots + (2n - 1)^2 = \dfrac{n(2n - 1)(2n + 1)}{3}$

11. P_n: $\dfrac{1}{1 \cdot 2} + \dfrac{1}{2 \cdot 3} + \cdots + \dfrac{1}{n \cdot (n + 1)} = \dfrac{n}{n + 1}$

12. P_n: $\displaystyle\sum_{k=1}^{n} k^4 = \frac{n(n + 1)(2n + 1)(3n^2 + 3n - 1)}{30}$

In Exercises 13–20, use mathematical induction to prove that the statement holds for all positive integers.

13. $1 + 5 + 9 + \cdots + (4n - 3) = n(2n - 1)$

14. $1 + 2 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$

15. $\dfrac{1}{1 \cdot 2} + \dfrac{1}{2 \cdot 3} + \dfrac{1}{3 \cdot 4} + \cdots + \dfrac{1}{n(n + 1)} = \dfrac{n}{n + 1}$

16. $\dfrac{1}{1 \cdot 3} + \dfrac{1}{3 \cdot 5} + \cdots + \dfrac{1}{(2n - 1)(2n + 1)} = \dfrac{n}{2n + 1}$

17. $2^n \geq 2n$

18. $3^n \geq 3n$

19. 3 is a factor of $n^3 + 2n$.

20. 6 is a factor of $7^n - 1$.

In Exercises 21 and 22, use *mathematical induction* to prove that the statement holds for all positive integers. (We have already seen each proved in another way.)

21. The sum of the first n terms of a geometric sequence with first term a_1 and common ratio $r \neq 1$ is $a_1(1 - r^n)/(1 - r)$.

22. The sum of the first n terms of an arithmetic sequence with first term a_1 and common difference d is

$$S_n = \frac{n}{2}[2a_1 + (n - 1)d].$$

In Exercises 23 and 24, use mathematical induction to prove that the formula holds for all positive integers.

23. Triangular Numbers $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$

24. Sum of the First n Cubes $\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$

[Note that if you put the results from Exercises 23 and 24 together, you obtain the pleasantly surprising equation

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = (1 + 2 + 3 + \cdots + n)^2.]$$

In Exercises 25–30, use the results of Exercises 21–24 and Example 2 to find the sums.

25. $1 + 2 + 3 + \cdots + 500$

26. $1^2 + 2^2 + \cdots + 250^2$

27. $4 + 5 + 6 + \cdots + n$

28. $1^3 + 2^3 + 3^3 + \cdots + 75^3$

29. $1 + 2 + 4 + 8 + \cdots + 2^{34}$

30. $1 + 8 + 27 + \cdots + 3375$

In Exercises 31–34, use the results of Exercises 21–24 and Example 2 to find the sum in terms of n.

31. $\sum_{k=1}^{n} (k^2 - 3k + 4)$

32. $\sum_{k=1}^{n} (2k^2 + 5k - 2)$

33. $\sum_{k=1}^{n} (k^3 - 1)$

34. $\sum_{k=1}^{n} (k^3 + 4k - 5)$

35. Group Activity Here is a proof by mathematical induction that any gathering of n people must all have the same blood type.

(Anchor) If there is 1 person in the gathering, everyone in the gathering obviously has the same blood type.

(Inductive Hypothesis) Assume that any gathering of k people must all have the same blood type.

(Inductive Step) Suppose $k + 1$ people are gathered. Send one of them out of the room. The remaining k people must all have the same blood type (by the inductive hypothesis). Now bring the first person back and send someone else out of the room. You get another gathering of k people, all of whom must have the same blood type. Therefore all $k + 1$ people must have the same blood type, and we are done by mathematical induction.

This result is obviously false, so there must be something wrong with the proof. Explain where the proof goes wrong.

36. Writing to Learn Kitty is having trouble understanding mathematical induction proofs because she does not understand the inductive hypothesis. If we can assume it is true for k, she asks, why can't we assume it is true for n and be done with it? After all, a variable is a variable! Write a response to Kitty to clear up her confusion.

Standardized Test Questions

37. True or False The goal of mathematical induction is to prove that a statement P_n is true for all real numbers n. Justify your answer.

38. True or False If P_n is the statement "$(n + 1)^2 = 4n$," then P_1 is true. Justify your answer.

You may use a graphing calculator when solving Exercises 39–42.

39. In a proof by mathematical induction that $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$ for all positive integers n, the inductive hypothesis would be to assume that

(a) $n = 1$.

(b) $k = 1$.

(c) $1 = \frac{1(1+1)}{2}$.

(d) $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$ for all positive integers n.

(e) $1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}$ for some positive integer k.

40. The first step in a proof by mathematical induction is to prove

(a) the anchor.

(b) the inductive hypothesis.

(c) the inductive step.

(d) the inductive principle.

(e) the inductive foundation.

41. Which of the following could be used to prove that $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ for all positive integers n?

I. Mathematical induction

II. The formula for the sum of a finite arithmetic sequence

III. The formula for the sum of a finite geometric sequence

(a) I only

(b) I and II only

(c) I and III only

(d) II and III only

(e) I, II, and III

42. Mathematical induction can be used to prove that, for any positive integer n, $\sum_{k=1}^{n} k^3 =$

(a) $\frac{n(n+1)}{2}$

(b) $\frac{n^2(n+1)^2}{2}$

(c) $\frac{n^2(n+1)^2}{4}$

(d) $\frac{n^3(n+1)^3}{2}$

(e) $\frac{n^3(n+1)^3}{8}$

Explorations

43. Use mathematical induction to prove that 2 is a factor of $(n+1)(n+2)$ for all positive integers n.

44. Use mathematical induction to prove that 6 is a factor of $n(n+1)(n+2)$ for all positive integers n. (You may assume the assertion in Exercise 43 to be true.)

45. Give an alternate proof of the assertion in Exercise 43 based on the fact that $(n+1)(n+2)$ is a product of two consecutive integers.

46. Give an alternate proof of the assertion in Exercise 44 based on the fact that $n(n+1)(n+2)$ is a product of three consecutive integers.

Extending the Ideas

In Exercises 47 and 48, use mathematical induction to prove that the statement holds for all positive integers.

47. **Fibonacci Sequence and Series** $F_{n+2} - 1 = \sum_{k=1}^{n} F_k$, where $\{F_n\}$ is the Fibonacci sequence.

48. If $\{a_n\}$ is the sequence $\sqrt{2}, \sqrt{2+\sqrt{2}}, \sqrt{2+\sqrt{2+\sqrt{2}}}, \ldots$, then $a_n < 2$.

49. Let a be any integer greater than 1. Use mathematical induction to prove that $a - 1$ divides $a^n - 1$ evenly for all positive integers n.

50. Give an alternate proof of the assertion in Exercise 49 based on the Factor Theorem of Section 2.4.

It is not necessary to anchor a mathematical induction proof at $n = 1$; we might only be interested in the integers greater than or equal to some integer c. In this case, we simply modify the anchor and inductive step as follows:

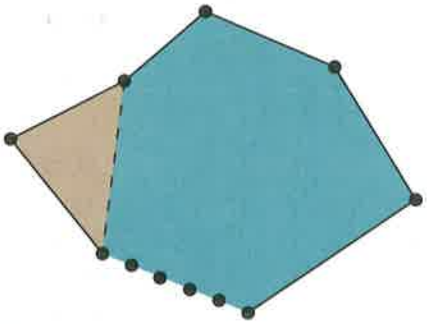
Anchor: P_c is true.

Inductive Step: If P_k is true for some $k \geq c$, then P_{k+1} is true.

(This is sometimes called the **Extended Principle of Mathematical Induction**.) Use this principle to prove the statements in Exercises 51 and 52.

51. $3n - 4 \geq n$, for all $n \geq 2$ **52.** $2^n \geq n^2$, for all $n \geq 4$

53. **Proving the interior angle formula** Use extended mathematical induction to prove P_n for $n \geq 3$.
P_n: The sum of the interior angles of an n-sided polygon is $180°(n-2)$.



9.6 STATISTICS AND DATA (GRAPHICAL)

What you'll learn about

- Statistics
- Displaying Categorical Data
- Stemplots
- Frequency Tables
- Histograms
- Time Plots

. . . and why

Graphical displays of data are increasingly prevalent in professional and popular media. We all need to understand them.

Statistics

Statistics is a branch of science that draws from both discrete mathematics and the mathematics of the continuum. The aim of statistics is to draw meaning from data and communicate it to others.

The objects described by a set of data are **individuals**, which may be people, animals, or things. The characteristic of the individuals being identified or measured is a **variable**. Variables are either *categorical* or *quantitative*. If the variable identifies each individual as belonging to a distinct class, such as male or female, then the variable is a **categorical variable**. If the variable takes on numerical values for the characteristic being measured, then the variable is a **quantitative variable**.

Examples of quantitative variables are heights of people, and weights of lobsters. You have already seen many tables of quantitative data in this book; indeed, most of the data-based exercises have been solved using techniques that are basic tools for statisticians. So far, however, most of our attention has been restricted to finding models that relate quantitative variables to each other. In the last two sections of this chapter, we will look at some of the other graphical and algebraic tools that can be used to draw meaning from data and communicate it to others.

Displaying Categorical Data

The National Center for Health Statistics reported that the leading causes of death in 1999 were heart disease, cancer, and stroke. Table 9.3 gives more detailed information.

TABLE 9.3 LEADING CAUSES OF DEATH IN THE UNITED STATES IN 1999

Cause of Death	Number of Deaths	Percentage
Heart Disease	725,192	30.3
Cancer	549,838	23.0
Stroke	167,366	7.0
Other	949,003	39.7

Source: National Vital Statistics Reports, Volume 49 Number 11, National Center for Health Statistics.

Because the causes of death are categories, not numbers, "cause of death" is a categorical variable. The numbers of deaths and percentages, while certainly numerical, are not values of a *variable*, because they do not describe the *individuals*. Nonetheless, the numbers can communicate information about the categorical variables by telling us the relative size of the categories in the 1999 population.

We can get that information directly from the numbers, but it is very helpful to see the comparative sizes visually. This is why you will often see categorical data displayed graphically, as a **bar chart** (Figure 9.13a), a **pie chart** (Figure 9.13b), or a **circle graph** (Figure 9.13c). For variety, the popular press also makes use of **picture graphs** suited to the categories being displayed. For example, the bars in Figure 9.13a could be made to look like tombstones of different sizes to emphasize that these are causes of death. In each case, the graph provides a visualization of the relative sizes of the categories, with the pie chart and circle graph providing the added visualization of how the categories are parts of a whole population.

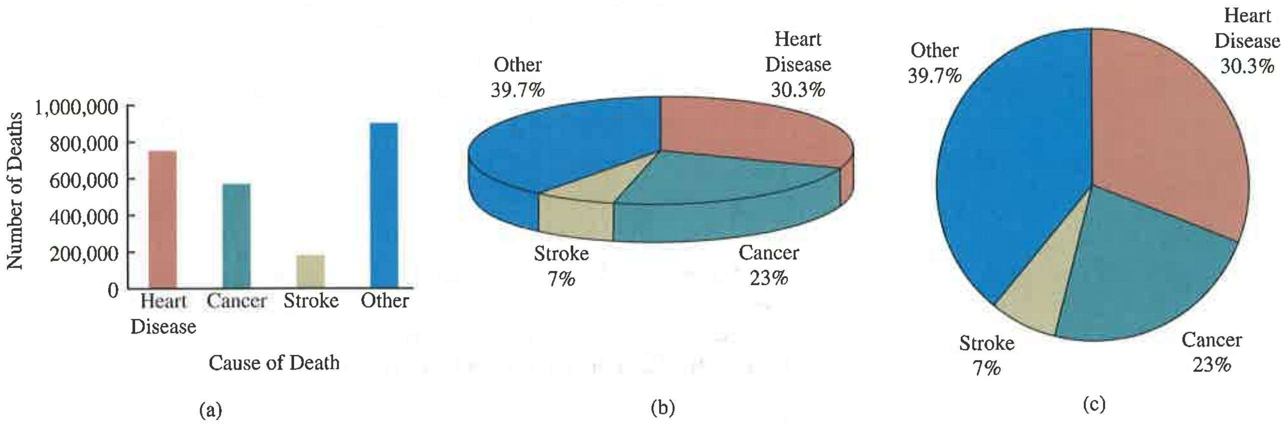


FIGURE 9.13 Causes of death in the United States in 1999 shown in (a) a bar graph, (b) a 3-D pie chart, and (c) a circle graph.

In bar charts of categorical data, the y-axis has a numerical scale, but the x-axis is labeled by category. The rectangular bars are separated by spaces to show that no continuous numerical scale is implied. (In this respect, a bar graph differs from a *histogram*, to be described later in this section.) A circle graph or a pie chart consists of shaded or colored sections of a circle or "pie." The central angles for the sectors are found by multiplying the percentages by 360°. For example, the angle for the sector representing cancer victims in Figure 9.13c is

$$23\% \cdot 360° = 82.8°.$$

It used to require time, skill, and mathematical savvy to draw data displays that were both visually appealing and geometrically accurate, but modern spreadsheet programs have made it possible for anyone with a computer to produce high-quality graphs from tabular data with the click of a button.

Stemplots

A quick way to organize and display a small set of quantitative data is with a **stemplot**, also called a **stem-and-leaf plot**. Each number in the data set is split into a **stem**, consisting of its initial digit or digits, and a **leaf**, which is its final digit.

EXAMPLE 1 Making a stemplot

Table 9.4 gives the percentage of each state's population in 2000 that was 65 or older. Make a stemplot for the data.

TABLE 9.4 PERCENTAGES OF STATE RESIDENTS IN 2000 WHO WERE 65 OR OLDER

AL	13.0	HI	13.3	MA	13.5	NM	11.7	SD	14.3
AK	5.7	ID	11.3	MI	12.3	NY	12.9	TN	12.4
AZ	13.0	IL	12.1	MN	12.1	NC	12.0	TX	9.9
AR	14.0	IN	12.4	MS	12.1	ND	14.7	UT	8.5
CA	10.6	IO	14.9	MO	13.5	OH	13.3	VT	12.7
CO	9.7	KS	13.3	MT	13.4	OK	13.2	VA	11.2
CT	13.8	KY	12.5	NE	13.6	OR	12.8	WA	11.2
DE	13.0	LA	11.6	NV	11.0	PA	15.6	WV	15.3
FL	17.6	ME	14.4	NH	12.0	RI	14.5	WI	13.1
GA	9.6	MD	11.3	NJ	13.2	SC	12.1	WY	11.7

Source: U.S. Census Bureau, 2001.

SOLUTION To form the stem-and-leaf plot, we use the whole number part of each number as the stem and the tenths digit as the leaf. We write the stems in order down the first column and, for each number, write the leaf in the appropriate stem row. Then we arrange the leaves in each stem row in ascending order. The final plot looks like this:

Stem	Leaf
5	7
6	
7	
8	5
9	6 7 9
10	6
11	0 2 2 3 3 6 7 7
12	0 0 1 1 1 1 3 4 4 5 7 8 9
13	0 0 0 1 2 2 3 3 3 4 5 5 6 8
14	0 3 4 5 7 9
15	3 6
16	
17	6

Notice that we include the "leafless" stems (6, 7, 16) in our plot, as those empty gaps are significant features of the visualization. For the same reason, be sure that each "leaf" takes up the same space along the stem. A branch with twice as many leaves should appear to be twice as long.

Now try Exercise 1.

EXPLORATION 1 Using Information from a Stemplot

By looking at both the stemplot and the table, answer the following questions about the distribution of senior citizens among the 50 states.

1. Judging from the stemplot, what was the approximate *average* national percentage of residents who were 65 or older?
2. In how many states were more than 15% of the residents 65 or older?
3. Which states were in the bottom tenth of all states in this statistic?
4. The numbers 5.7 and 17.6 are so far above or below the other numbers in this stemplot that statisticians would call them *outliers*. Quite often there is some special circumstance that sets outliers apart from the other individuals under study and explains the unusual data. What could explain the two outliers in this stemplot?

Sometimes the data are so tightly clustered that a stemplot has too few stems to give a meaningful visualization of the data. In such cases we can spread the data out by splitting the stems, as in Example 2.

EXAMPLE 2 Making a split-stem stemplot

The average annual salaries for the top 15 U.S. metropolitan areas (in this category) are shown in Table 9.5. Make a stemplot that provides a good visualization of the data. What is the average of the 15 numbers? Why is the stemplot a better summary of the data than the average?

SOLUTION We first round the data to $1000 units, which does not affect the visualization. Then, to spread out the data a bit, we *split* each stem, putting leaves 0–4 on the lower stem and leaves 5–9 on the upper stem:

Stem	Leaf
4	2 4 4
4	5 5 5 5 6 8 9 9
5	1
5	6 9
6	
6	
7	
7	6

The average of the 15 numbers is $49,582, but this is misleading. The stemplot shows that 11 of the numbers are lower than that, while only four are higher. It is better to observe that the distribution is clustered fairly tightly around $45,000, with the San Francisco and New York salaries slightly higher and the San Jose salary a significant outlier on the high end.

Now try Exercise 3.

TABLE 9.5 THE AVERAGE ANNUAL SALARIES (IN DOLLARS) FOR THE TOP 15 U.S. METROPOLITAN AREAS

San Jose, CA	76,076	Newark, NJ	48,733	Seattle, WA[5]	45,171
San Francisco, CA	59,314	Jersey City, NJ	47,514	Trenton, NJ	44,576
New York, NY	56,377	Boulder/Longmont, CO	45,565	Oakland, CA	44,170
New Haven, CT[1]	50,585	Washington, D.C.[3]	45,333	Bergen/Passaic, NJ	43,789
Middlesex, NJ[2]	48,977	Boston, MA[4]	45,191	Hartford, CT	42,349

[1] Includes Bridgeport, Stamford, Danbury, Waterbury

[2] Includes Somerset, Hunterdon

[3] Includes suburbs in MD, VA, WV

[4] Includes Worcester, Lawrence, Lowell, Brockton, and some NH

[5] Includes Bellevue and Everett

Source: Bureau of Labor Statistics as quoted by World Almanac and Book of Facts, 2002.

Sometimes it is easier to compare two sets of data if we have a visualization that allows us to view both stemplots simultaneously. **Back-to-back stemplots** use the same stems, but leaves from one set of data are added to the left, while leaves from another set are added to the right.

EXAMPLE 3 Making back-to-back stemplots

Mark McGwire and Barry Bonds entered the major leagues in 1986. From 1986 to 2001, they averaged 36.44 and 35.44 home runs per year, respectively. Compare their annual home run totals with a back-to-back stemplot. Can you tell which player has been more consistent as a home run hitter?

SOLUTION We form a back-to-back stemplot with McGwire's totals branching off to the left and Bond's to the right.

Mark McGwire		Barry Bonds
9 9 3	0	
	1	6 9
9 2	2	4 5 5
9 9 3 2 2	3	3 3 4 4 7 7
9 2	4	0 2 6 9
8 2	5	
5	6	
0	7	3

The single-digit home run years for McGwire can be explained by fewer times at bat (late entry into the league in 1986 and injuries in 1993 and 1994). If those years are ignored as anomalies, McGwire's numbers seem to indicate more consistency. Bond's record-setting 73 in 2001 was an outlier, although he continued that pace in 2002.

Now try Exercise 5.

Table 9.6 Major League Home Run Totals for Mark McGwire and Barry Bonds through 2001

Year	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01
McGwire	3	49	32	33	39	22	42	9	9	39	52	58	70	65	32	29
Bonds	16	25	24	19	33	25	34	46	37	33	42	40	37	34	49	73

Source: Major League Baseball Enterprises, 2002.

Frequency Tables

The visual impact of a stemplot comes from the lengths of the various rows of leaves, which is just a way of seeing *how many* leaves branch off each stem. The number of leaves for a particular stem is the **frequency** of observations within each stem interval. Frequencies are often recorded in a **frequency table**. Table 9.7 shows a frequency table for Mark McGwire's yearly home run totals from 1986 to 2001 (see Example 3). The table shows the **frequency distribution**—literally the way that the total frequency of 16 is "distributed" among the various home run intervals. This same information is conveyed visually in a stemplot, but notice that the stemplot has the added numerical advantage of displaying what the numbers in each interval actually are.

Table 9.7 Frequency Table for Mark McGwire's Yearly Home Run Totals, 1986–2001 (Higher frequencies in a table correspond to longer leaf rows in a stemplot. Unlike a stemplot, a frequency table does not display what the numbers in each interval actually are.)

Home Runs	Frequency	Home Runs	Frequency
0–9	3	40–49	2
10–19	0	50–59	2
20–29	2	60–69	1
30–39	5	70–79	1
		Total	16

Histograms

A **histogram**, closely related to a stemplot, displays the information of a frequency table. Visually, a histogram is to quantitative data what a bar chart is to categorical data. Unlike a bar chart, however, both axes of a histogram have numerical scales, and the rectangular bars on adjacent intervals have no gaps between them.

Figure 9.14 shows a histogram of the information in Table 9.7, where each bar corresponds to an interval in the table and the height of each bar represents the frequency of observations within the interval.

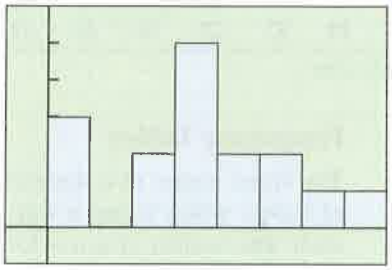


FIGURE 9.14 A histogram showing the distribution of Mark McGwire's annual home run totals from 1986 to 2001. This is a visualization of the data in Table 9.7.

EXAMPLE 4 Graphing a histogram on a calculator

Make a histogram of Hank Aaron's annual home run totals given in Table 9.8, using intervals of width 5.

TABLE 9.8 REGULAR SEASON HOME RUN STATISTICS FOR HANK AARON

Year	Home Runs	Year	Home Runs	Year	Home Runs
1954	13	1962	45	1970	38
1955	27	1963	44	1971	47
1956	26	1964	24	1972	34
1957	44	1965	32	1973	40
1958	30	1966	44	1974	20
1959	39	1967	39	1975	12
1960	40	1968	29	1976	10
1961	34	1969	44		

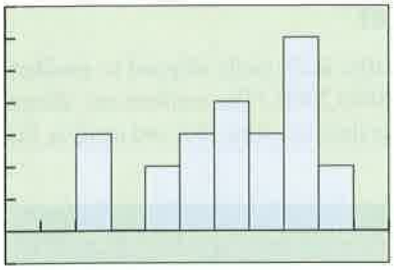
Source: The Baseball Encyclopedia (7th ed., 1988, New York: MacMillan) p. 695.

SOLUTION We first make a frequency table for the data, using intervals of width 5. (This is not needed for the calculator to produce the histogram, but we will compare this with the result.)

Home Runs	Frequency	Home Runs	Frequency
10–14	3	30–34	4
15–19	0	35–39	3
20–24	2	40–44	6
25–29	3	45–49	2
		Total	23

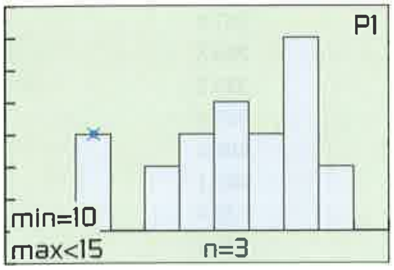
To scale the x-axis to be consistent with the intervals of the table, let Xmin = 0, Xmax = 55, and Xscl = 5. Notice that the maximum frequency is 6 years (40–44 home runs), so the y-axis ought to extend at least to 7. Enter the data from Table 9.10 into list L1 and plot a histogram in the window $[0, 55]$ by $[-1, 7]$. (See Figure 9.15a.) Tracing along the histogram should reveal the same frequencies as in the frequency table we made. (See Figure 9.15b.)

Now try Exercise 11.



[0, 55] by [−1, 7]

(a)



[0, 55] by [−1, 7]

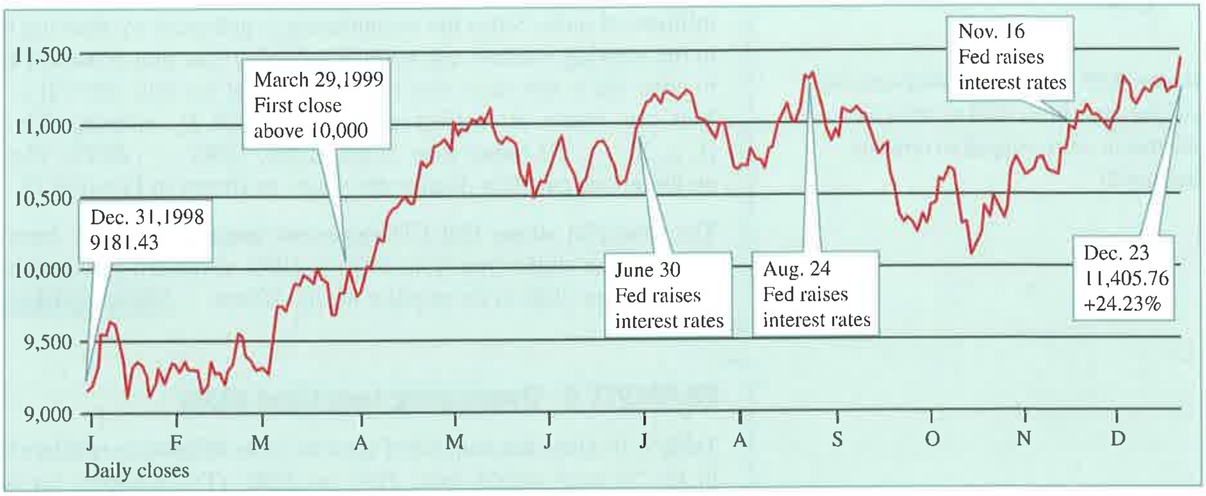
(b)

FIGURE 9.15 Calculator histograms of Hank Aaron's yearly home run totals. (Example 4)

Time Plots

We have seen in this book many examples of functions in which the input variable is time. It is also quite common to consider quantitative data as a function of time. If we make a scatter plot of the data (y) against the time (x) that it was measured, we can analyze the patterns as the variable changes over time. To help with the visualization, the discrete points from left to right are connected by line segments, just as a grapher would do in connect mode. The resulting **line graph** is called a **time plot**.

Time plots reveal trends in data over time. These plots frequently appear in magazines and newspapers and on the Internet, a typical example being the graph of the historic 1999 climb of the Dow Jones Industrial Average (DJIA) in Figure 9.16.



Dec. 31, 1998
9181.43

March 29, 1999
First close
above 10,000

June 30
Fed raises
interest rates

Aug. 24
Fed raises
interest rates

Nov. 16
Fed raises
interest rates

Dec. 23
11,405.76
+24.23%

FIGURE 9.16 Time plot of the Dow Jones Industrial Average during the spectacular year 1999. Investors get a good visualization of where the stock market has been, although the trick is to figure out where it is going. *(Source: Quote.com, as reported by the Associated Press in the Chattanooga Times/Free Press.)*

EXAMPLE 5 Drawing a time plot

Table 9.9 gives the number of compact disc (CD) units shipped to retailers each year in the 13-year period from 1988 to 2000. (The numbers are shown in millions, net after returns.) Display the data in a time plot and analyze the 13-year trend.

TABLE 9.9 THE NUMBER OF CDS SHIPPED TO RETAILERS EACH YEAR FROM 1988 TO 2000, IN MILLIONS OF UNITS

Year	CDs (millions)
1988	149.7
1989	207.2
1990	286.5
1991	333.3
1992	407.5
1993	495.4
1994	662.1
1995	722.9
1996	778.9
1997	753.1
1998	847.0
1999	938.9
2000	942.5

Source: Recording Industry Association of America as reported in the World Almanac and Book of Facts, 2002.

SOLUTION The horizontal axis represents time (in years) from 1988 to 2000. The vertical axis represents the number of CDs shipped that year, in millions of units. Since the visualization is enhanced by showing both axes in the viewing window, the vertical axis of a time plot is usually translated to cross the x-axis at or near the beginning of the time interval in the data. You can create this effect on your grapher by entering the years as $\{1, 2, 3, \ldots, 13\}$ rather than $\{1988, 1989, 1990, \ldots, 2000\}$. The labeling on the x-axis can then display the years, as shown in Figure 9.17.

The time plot shows that CD shipments increased steadily from 1988 to 2000, with a slight drop from 1996 to 1997, attributed by some analysts to a sudden age shift in the popular music culture. Now try Exercise 17.

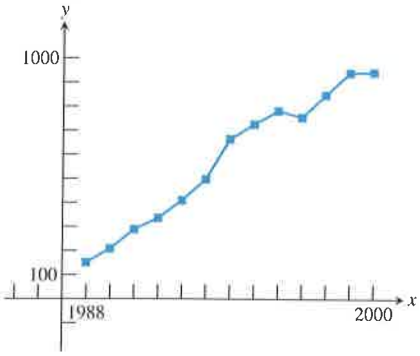


FIGURE 9.17 A time plot of CD volume over the years from 1988 to 2000, as reflected in units shipped to retailers. (Example 5)

EXAMPLE 6 Overlaying two time plots

Table 9.10 gives the number of cassette units shipped to retailers each year in the 13-year period from 1988 to 2000. (The numbers are shown in millions, net after returns.) Compare the cassette trend with the CD trend by overlaying the time plots for the two products.

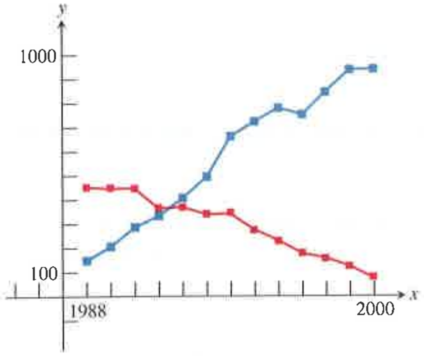


FIGURE 9.18 A time plot comparing cassette shipments and CD shipments for the time period from 1988 to 2000. (Example 6)

TABLE 9.10 THE NUMBER OF CASSETTES SHIPPED TO RETAILERS EACH YEAR FROM 1988 TO 2000, IN MILLIONS OF UNITS

Year	Cassettes (millions)
1988	450.1
1989	446.2
1990	442.2
1991	360.1
1992	366.4
1993	339.5
1994	345.4
1995	272.6
1996	225.3
1997	172.6
1998	158.5
1999	123.6
2000	76.0

Source: Recording Industry Association of America as reported in the World Almanac and Book of Facts, 2002.

SOLUTION The two time plots are shown in Figure 9.18. The popularity of cassette tapes declined as more and more consumers switched to CD technology. The CD spurt in 1994 foreshadowed a steeper decline in cassette shipments thereafter. Now try Exercise 19.

QUICK REVIEW 9.6

In Exercises 1–6, solve for the required value.

1. 457 is what percent of 2953?
2. 827 is what percent of 3950?
3. 52° is what percent of 360°?
4. 98° is what percent of 360°?
5. 734 is 42.6% of what number?
6. 5106 is 55.5% of what number?

In Exercises 7–10, round the given value to the nearest whole number in the specified units.

7. $234,598.43 (thousands of dollars)
8. 237,834,289 (millions)
9. 848.36 thousands (millions)
10. 1432 millions (billions)

SECTION 9.6 EXERCISES

Table 9.11 shows the home run statistics for Roger Maris during his major league career.

TABLE 9.11 REGULAR SEASON HOME RUN STATISTICS FOR ROGER MARIS

Year	Home Runs
1957	14
1958	28
1959	16
1960	39
1961	61
1962	33
1963	23
1964	26
1965	8
1966	13
1967	9
1968	5

Source: The Baseball Encyclopedia, 7th ed., 1988.

1. Make a stemplot of the data in Table 9.11. Are there any outliers?
2. Make a back-to-back stemplot comparing the annual home run production of Roger Maris (Table 9.11) with that of Hank Aaron (Table 9.8 on page 764). Write a brief interpretation of the stemplot.

In Exercises 3–6, construct the indicated stemplot from the data in Table 9.12. Then write a brief interpretation of the stemplot.

TABLE 9.12 LIFE EXPECTANCY BY GENDER FOR THE NATIONS OF SOUTH AMERICA

Nation	Male	Female
Argentina	71.9	78.8
Bolivia	61.5	66.7
Brazil	59.0	67.7
Chile	72.6	79.4
Colombia	66.7	74.6
Ecuador	68.5	74.3
Guyana	60.5	66.2
Paraguay	71.4	76.5
Peru	67.9	72.8
Suriname	69.0	74.4
Uruguay	72.1	79.0
Venezuela	70.3	76.6

Source: The World Almanac and Book of Facts, 2002.

3. A stemplot showing life expectancies of males in the nations of South America (Round to nearest year and use split stems.)
4. A stemplot showing life expectancies of females in the nations of South America (Round to nearest year and use split stems.)
5. A back-to-back stemplot for life expectancies of males and females in the nations of South America (Round to nearest year and use split stems.)
6. A stemplot showing the difference between female and male life expectancies in the nations of South America (Use unrounded data and do not split stems.)

In Exercises 7 and 8, use the data in Table 9.12 to construct the indicated frequency table, using intervals 55.0–59.9, 60.0–64.9, 65.0–69.9, etc.

7. Life expectancies of males in the nations of South America
8. Life expectancies of females in the nations of South America

In Exercises 9–12, draw a histogram for the given table.

9. The frequency table in Exercise 7
10. The frequency table in Exercise 8
11. Table 9.13 of Willie Mays' annual home run totals, using intervals 1–5, 6–10, 11–15, etc.
12. Table 9.13 of Mickey Mantle's annual home run totals, using intervals 1–5, 6–10, 11–15, etc.

TABLE 9.13 REGULAR SEASON HOME RUN STATISTICS FOR WILLIE MAYS AND MICKEY MANTLE

Year	Mays	Mantle	Year	Mays	Mantle
1951	20	13	1962	38	30
1952	4	23	1963	47	15
1953	41	21	1964	52	35
1954	51	27	1965	37	19
1955	36	37	1966	22	23
1956	35	52	1967	23	22
1957	29	34	1968	13	18
1958	34	42	1969	28	
1959	29	31	1970	18	
1960	40	40	1971	8	
1961	49	54	1972	6	

Source: The Baseball Encyclopedia, 7th ed., 1988.

In Exercises 13–16, make a time plot for the indicated data.

13. Willie Mays's annual home run totals given in Table 9.13

14. Mickey Mantle's annual home run totals given in Table 9.13

15. Mark McGwire's home run totals given in Table 9.6 on page 763

16. Hank Aaron's home run totals given in Table 9.8 on page 764

Table 9.14 shows the total amount of money won (in units of \$1000, rounded to the nearest whole number) by the leading money winners in women's (LPGA) and men's (PGA) professional golf for selected years between 1955 and 1997.

TABLE 9.14 YEARLY EARNINGS (IN THOUSANDS OF DOLLARS) OF THE TOP MONEY WINNERS IN MEN'S AND WOMEN'S GOLF FOR SELECTED YEARS FROM 1955 TO 2000 (ROUNDED TO THE NEAREST \$1000)

Year	Men (PGA)	Women (LPGA)
1955	65	16
1960	75	17
1965	141	29
1970	157	30
1975	323	95
1980	531	231
1985	542	416
1989	1395	654
1992	1344	693
1995	1655	667
1996	1780	1002
1997	2067	1237
1998	2591	1093
1999	6617	1592
2000	9188	1877

Source: The World Almanac and Book of Facts, 2002.

17. Make a time plot for the men's winnings in Table 9.14. Write a brief interpretation of the time plot.

18. Make a time plot for the women's winnings in Table 9.14. Write a brief interpretation of the time plot.

19. Compare the trends in Table 9.14 by overlaying the time plots. Write a brief interpretation.

20. **Writing to Learn** (Continuation of Exercise 19) The data in Table 9.14 show that the earnings for the top PGA player rose by a modest 68% in the decade from 1975 to 1985, while the earnings for the top LPGA player rose by a whopping 338%. Although this was, in fact, a strong growth period for women's sports, statisticians would be unlikely to draw any conclusions from a comparison of these two numbers. Use the visualization from the comparative time plot in Exercise 19 to explain why.

In Exercises 21 and 22, compare performances by overlaying time plots.

21. The time plots from Exercises 13 and 14 to compare the performances of Mays and Mantle

22. The time plots from Exercises 15 and 16 to compare the performances of McGwire and Aaron

In Exercises 23 and 24, analyze the data as indicated.

23. The salaries of the workers in one department of the Garcia Brothers Company (given in thousands of dollars) are as follows:

33.5, 35.3, 33.8, 29.3, 36.7, 32.8, 31.7, 36.3, 33.5, 28.2, 34.8, 33.5, 35.3, 29.7, 38.5, 32.7, 34.8, 34.2, 31.6, 35.4

(a) Complete a stemplot for this data set.

(b) Create a frequency table for the data.

(c) Draw a histogram for the data. What viewing window did you use?

(d) Why does a time plot not work well for the data?

24. The average wind speeds for one year at 44 climatic data centers around the United States are as follows:

9.0, 6.9, 9.1, 9.2, 10.2, 12.5, 12.0, 11.2, 12.9, 10.3, 10.6, 10.9, 8.7, 10.3, 11.0, 7.7, 11.4, 7.9, 9.6, 8.0, 10.7, 9.3, 7.9, 6.2, 8.3, 8.9, 9.3, 11.6, 10.6, 9.0, 8.2, 9.4, 10.6, 9.5, 6.3, 9.1, 7.9, 9.7, 8.8, 6.9, 8.7, 9.0, 8.9, 9.3

(a) Complete a stemplot for this data set.

(b) Create a frequency table for the data.

(c) Draw a histogram for the data. What viewing window did you use?

(d) Why does a circle graph not work well for the data?

In Exercises 25 and 26, compare by overlaying time plots for the data in Table 9.15

TABLE 9.15 POPULATION (IN MILLIONS) OF CALIFORNIA, FLORIDA, ILLINOIS, NEW YORK, PENNSYLVANIA, AND TEXAS

Year	CA	FL	IL	NY	PA	TX
1900	1.5	0.5	4.8	7.3	6.3	3.0
1910	2.4	0.8	5.6	9.1	7.7	3.9
1920	3.4	1.0	6.5	10.4	8.7	4.7
1930	5.7	1.5	7.6	12.6	9.6	5.8
1940	6.9	1.9	7.9	13.5	9.9	6.4
1950	10.6	2.7	8.7	14.8	10.5	7.7
1960	15.7	5.0	10.0	16.8	11.3	9.6
1970	20.0	6.8	11.1	18.2	11.8	11.2
1980	23.7	9.7	11.4	17.6	11.9	14.2
1990	29.8	12.9	11.4	18.0	11.9	17.0

Source: U.S. Census Bureau, as reported in The World Almanac and Books of Facts, 2002.

25. The populations of California, New York, and Texas from 1900 through 1990

26. The populations of Florida, Illinois, and Pennsylvania from 1900 through 1990

Standardized Test Questions

27. **True or False** If there are stems without leaves in the interior of a stemplot, it is best to omit them. Justify your answer.

28. **True or False** The highest and lowest numbers in a set of data are called outliers. Justify your answer.

Answer Exercises 29–32 without using a calculator.

29. A time plot is an example of a

(a) histogram.

(b) bar graph.

(c) line graph.

(d) pie chart.

(e) table.

30. A back-to-back stemplot is particularly useful for

(a) identifying outliers.

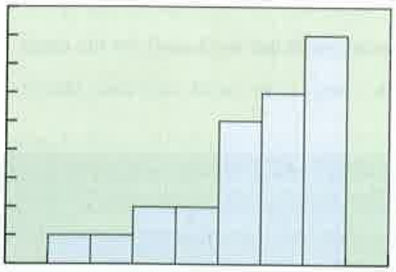
(b) comparing two data distributions.

(c) merging two sets of data.

(d) graphing home runs.

(e) distinguishing stems from leaves.

31. The histogram below would most likely result from which set of data?



(a) test scores from a fairly easy test

(b) weights of children in a third-grade class

(c) winning soccer scores for a team over the course of a season

(d) ages of all the people visiting the Bronx Zoo at a given point in time

(e) prices of all the desserts on the menu at a certain restaurant

32. A sector on a pie chart with a central angle of 45° corresponds to what percentage of the data?

(a) 8%

(b) 12.5%

(c) 15%

(d) 25%

(e) 45%

Explorations

33. **Group Activity** Measure the resting pulse rates (beats per minute) of the members of your class. Make a stemplot for the data. Are there any outliers? Can they be explained?

34. **Group Activity** Measure the heights (in inches) of the members of your class. Make a back-to-back stemplot to compare the distributions of the male heights and the female heights. Write a brief interpretation of the stemplot.

Extending the Ideas

35. **Time Plot of Periodic Data** Some data are periodic functions of time. If data vary in an annual cycle, the period is 1 year. Use the information in Table 9.16 to overlay the time plots for the average daily high and low temperatures for Beijing, China.

Table 9.16 Average Daily High and Low Temperatures in °C for Beijing, China

Month	High	Low
January	2	−9
February	5	−7
March	12	−1
April	20	7
May	27	13
June	31	18
July	32	22
August	31	21
September	27	14
October	21	7
November	10	−1
December	3	−7

Source: National Geographic Atlas of the World (rev. 6th ed., 1992, Washington, D.C.), plate 132.

36. Find a sinusoidal function that models each time plot in Exercise 35. (See Sections 4.4 and 4.8.)

9.7 STATISTICS AND DATA (ALGEBRAIC)

What you'll learn about

- Parameters and Statistics
- Mean, Median, and Mode
- The Five-Number Summary
- Boxplots
- Variance and Standard Deviation
- Normal Distributions

. . . and why

The language of statistics is becoming more commonplace in our everyday world.

Parameters and Statistics

The various numbers that are associated with a data set are called **statistics**. They serve to describe the individuals from which the data come, so the gathering and processing of such numerical information is often called **descriptive statistics**. You saw many examples of descriptive statistics in Section 9.6.

The *science* of statistics comes in when we use descriptive statistics (like the results of a study of 1500 smokers) to make judgments, called *inferences*, about entire *populations* (like all smokers). Statisticians are really interested in the numbers called **parameters** that are associated with entire populations. Since it is usually either impractical or impossible to measure entire populations, statisticians gather statistics from carefully chosen **samples**, then use the science of **inferential statistics** to make inferences about the parameters.

EXAMPLE 1 Distinguishing a parameter from a statistic

A 1996 study called *Kids These Days: What Americans Really Think About the Next Generation* reported that 33% of adolescents say there is no adult at home when they return from school. The report was based on a survey of 600 randomly selected people aged 12 to 17 years old and had a margin of error of ±4% (Source: *Public Agenda*). Did the survey measure a parameter or a statistic, and what does that "margin of error" mean?

SOLUTION The survey did not measure all adolescents in the population, so it did not measure a parameter. They *sampled* 600 adolescents and found a statistic. On the other hand, the statement "33% of adolescents say there is no adult at home" is making an inference about *all* American adolescents. We should interpret that statement in terms of the margin of error, as meaning "between 29% and 37% of all American adolescents would say that there is no adult at home when they return from school." In other words, the statisticians are confident that the parameter is within ±4% of their sample statistic, even though they only sampled 600 adolescents—a tiny fraction of the adolescent population! The mathematics that gives them that confidence is based on the laws of probability and is scientifically reliable, but we will not go into it here.

Now try Exercise 1.

Mean, Median, and Mode

If you wanted to study the effect of chicken feed additives on the thickness of egg shells, you would need to sample many eggs from different hens under various feeding conditions. Suppose you were to gather data from 50 eggs from hens eating feed A and 50 eggs from hens eating feed B. How would you compare the two? The simplest way would be to find the *average* egg shell thickness for each feed and compare those two numbers.

The word "average," however, can have several different meanings, all of them somehow *measures of center*.

- If we say, "The average on last week's test was 83.4," we are referring to the **mean**. (This is what most people usually think of as "average.")
- If we say, "The average test score puts you right in the middle of the class," we are referring to the **median**.
- If we say, "The average American student starts college at age 18," we are referring to the **mode**.

We will look at each of these measures separately.

Definition Mean

The **mean** of a list of n numbers $\{x_1, x_2, \ldots, x_n\}$ is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The mean is also called the *arithmetic mean*, *arithmetic average*, or *average value*.

EXAMPLE 2 Computing a mean

Find the mean annual home run total for Roger Maris's major league career, 1957–1968 (Table 9.11 on page 768).

SOLUTION According to Table 9.11, we are looking for the mean of the following list of 12 numbers: {14, 28, 16, 39, 61, 33, 23, 26, 8, 13, 9, 5}.

$$\bar{x} = \frac{14 + 28 + 16 + \cdots + 9 + 5}{12} = \frac{275}{12} \approx 22.9$$

Now try Exercise 9.

As common as it is to use the mean as a measure of center, sometimes it can be misleading. For example, if you were to find the mean annual salary of a Geography major from the University of North Carolina working in Chicago in 1997, it would probably be a number in the millions of dollars. This is because the group being measured, which is not very large, includes an outlier named Michael Jordan. The mean can be strongly affected by outliers.

We call a statistic **resistant** if it is not strongly affected by outliers (See Exploration 1, Section 9.6). While the mean is not a resistant measure of center, the *median* is.

Definition Median

The **median** of a list of n numbers $\{x_1, x_2, \ldots, x_n\}$ arranged in order (either ascending or descending) is

- the middle number if n is odd, and
- the mean of the two middle numbers if n is even.

EXAMPLE 3 Finding a median

Find the median of Roger Maris's annual home run totals. (See Example 2.)

SOLUTION First, we arrange the list in ascending order: {5, 8, 9, 13, 14, 16, 23, 26, 28, 33, 39, 61}. Since there are 12 numbers, the median is the mean of the 6th and 7th numbers:

$$\frac{16 + 23}{2} = 19.5.$$

Notice that this number is quite a bit smaller than the mean (22.9). The mean is strongly affected by the outlier representing Maris's record-breaking season, while the median is not. The median would still have been 19.5 if he had hit only 41 home runs that season, or, indeed, if he had hit 81.

Now try Exercise 11.

Both the mean and the median are important measures of center. A somewhat less important measure of center (but a measure of possible significance statistically) is the *mode*.

À LA MODE

The mode can also be used for categorical variables.

Definition Mode

The **mode** of a list of numbers is the number that appears most frequently in the list.

EXAMPLE 4 Finding a mode

Find the mode for the annual home run totals of Hank Aaron (Table 9.8 on page 764). Is this number of any significance?

SOLUTION It helps to arrange the list in ascending order: {10, 12, 13, 20, 24, 26, 27, 29, 30, 32, 34, 34, 38, 39, 39, 40, 40, 44, 44, 44, 44, 45, 47}.

Most numbers in the list appear only once. Three numbers appear twice, and the number 44 appears four times. The mode is 44.

It is rather unusual to have that many repeats of a number in a list of this sort. (By comparison, the Maris list has no repeats—and hence no mode—while the lists for Mays, Mantle, McGwire, and Bonds contain no number more than twice.) The mode in this case has special significance only to baseball trivia buffs, who recognize 44 as Aaron's uniform number!

Now try Exercise 17.

Example 4 demonstrates why the mode is less useful as a measure of center. The mode (44) is a long way from the median (34) and the mean (32.83), either of which does a much better job of representing Aaron's annual home run output over the course of his career.

EXAMPLE 5 Using a frequency table

A teacher gives a 10-point quiz and records the scores in a frequency table (Table 9.17) as shown below. Find the mode, median, and mean of the data.

TABLE 9.17 QUIZ SCORES FOR EXAMPLE 5

Score	10	9	8	7	6	5	4	3	2	1	0
Frequency	2	2	3	8	4	3	3	2	1	1	1

SOLUTION The total of the frequencies is 30, so there are 30 scores.

The *mode* is 7, since that is the score with the highest frequency.

The *median* of 30 numbers will be the mean of the 15th and 16th numbers. The table is already arranged in descending order, so we count the frequencies from left to right until we come to 15. We see that the 15th number is a 7 and the 16th number is a 6. The median, therefore, is 6.5.

To find the *mean*, we multiply each number by its frequency, add the products, and divide the total by 30:

$$\bar{x} = \frac{\begin{array}{c}[10(2) + 9(2) + 8(3) + 7(8) + 6(4) + 5(3)\\ + 4(3) + 3(2) + 2(1) + 1(1) + 0(1)]\end{array}}{30}$$

$$= 5.9\overline{3}$$

Now try Exercise 19.

The formula for finding the mean of a list of numbers $\{x_1, x_2, \ldots, x_n\}$ with frequencies $\{f_1, f_2, \ldots, f_n\}$ is

$$\bar{x} = \frac{x_1f_1 + x_2f_2 + \cdots + x_nf_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum x_if_i}{\sum f_i}.$$

This same formula can be used to find a **weighted mean**, in which numbers $\{x_1, x_2, \ldots, x_n\}$ are given **weights** before the mean is computed. The weights act the same way as frequencies.

EXAMPLE 6 Working with a weighted mean

At Marty's school, it is an administrative policy that the final exam must count 25% of the semester grade. If Marty has an 88.5 average going into the final exam, what is the minimum exam score needed to earn a 90 for the semester?

SOLUTION The preliminary average (88.5) is given a weight of 0.75 and the final exam (x) is given a weight of 0.25. We will assume that a semester average of 89.5 will be rounded to a 90 on the transcript. Therefore,

$$\frac{88.5(0.75) + x(0.25)}{0.75 + 0.25} = 89.5$$

$$0.25x = 89.5(1) - 88.5(0.75)$$

$$x = 92.5$$

Interpreting the answer, we conclude that Marty needs to make a 93 on the final exam. Now try Exercise 21.

The Five-Number Summary

Measures of center tell only part of the story of a data set. They do *not* indicate how widely distributed or highly variable the data are. *Measures of spread* do. The simplest and crudest measure of spread is the **range**, which is the difference between the maximum and minimum values in the data set:

$$\text{Range} = \text{maximum} - \text{minimum}.$$

For example, the range of numbers in Roger Maris's annual home run production is $61 - 5 = 56$. Like the mean, the range is a statistic that is strongly influenced by outliers, so it can be misleading. A more resistant (and therefore more useful) measure is the *interquartile range*, which is the range of the middle half of the data.

Just as the median separates the data into halves, the **quartiles** separate the data into fourths. The **first quartile** Q_1 is the median of the lower half of the data, the **second quartile** is the median, and the **third quartile** Q_3 is the median of the upper half of the data. The **interquartile range** (IQR) measures the spread between the first and third quartiles, comprising the middle half of the data:

$$IQR = Q_3 - Q_1.$$

Taken together, the maximum, the minimum, and the three quartiles give a fairly complete picture of both the center and the spread of a data set.

FINDING QUARTILES

When we are finding the quartiles for a data set with an odd number of values, we do *not* consider the middle value to be included in either the lower or the upper half of the data.

Definition Five-Number Summary

The **five-number summary** of a data set is the collection

$$\{\text{minimum}, Q_1, \text{median}, Q_3, \text{maximum}\}.$$

COMPUTING STATISTICS ON A CALCULATOR

Modern calculators will usually process lists of data and give statistics like the mean, median, and quartiles with a push of a button. Consult your owner's manual.

EXAMPLE 7 Five-number summary and spread

Find the five-number summaries for the male and female life expectancies in South American nations (Table 9.12 on page 768 and compare the spreads.

SOLUTION Here are the lists in ascending order.

Males:

{59.0, 60.5, 61.5,   66.7, 67.9, 68.5,   69.0, 70.3, 71.4,   71.9, 72.1, 72.6}

Females:

{66.2, 66.7, 67.7,   72.8, 74.3, 74.4,   74.6, 76.5, 76.6,   78.8, 79.0, 79.4}

We have spaced the lists to show where the quartiles appear. The median of the 12 values is midway between the 6th and 7th values. The first quartile is the median of the lower 6 values (i.e., midway between the 3rd and 4th), and the third quartile is the median of the upper 6 values (i.e., midway between the 9th and 10th).

The five-number summaries are shown below.

Males: {59.0, 64.1, 68.75, 71.65, 72.6}

Females: {66.2, 70.25, 74.5, 77.7, 79.4}

The males have a range of $72.6 - 59.0 = 13.6$ and an *IQR* of $71.65 - 64.1 = 7.55$.

The females have a range of $79.4 - 66.2 = 13.2$ and an *IQR* of $77.7 - 70.25 = 7.45$.

Not only do the women live longer, but there is less variability in their life expectancies (as measured by the *IQR*). Male life expectancy is more strongly affected by different political conditions within countries (war, civil strife, crime, etc.).

Now try Exercise 23.

You can learn a lot about the data by considering the *shape* of the distribution, as visualized in a histogram. Try answering the questions in Exploration 1.

EXPLORATION 1 Interpreting Histograms

1. Of the two histograms shown in Figure 9.19, which displays a data set with more variability?
2. Of the three histograms in Figure 9.20, which has a median less than its mean? Which has a median greater than its mean? Which has a median approximately equal to its mean?

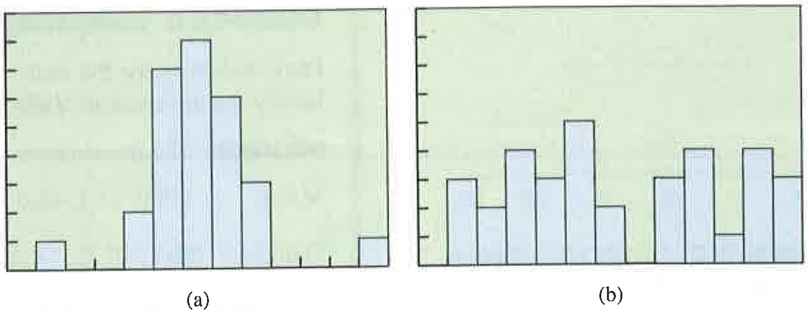


FIGURE 9.19 Which data set has more variability? (Exploration 1)

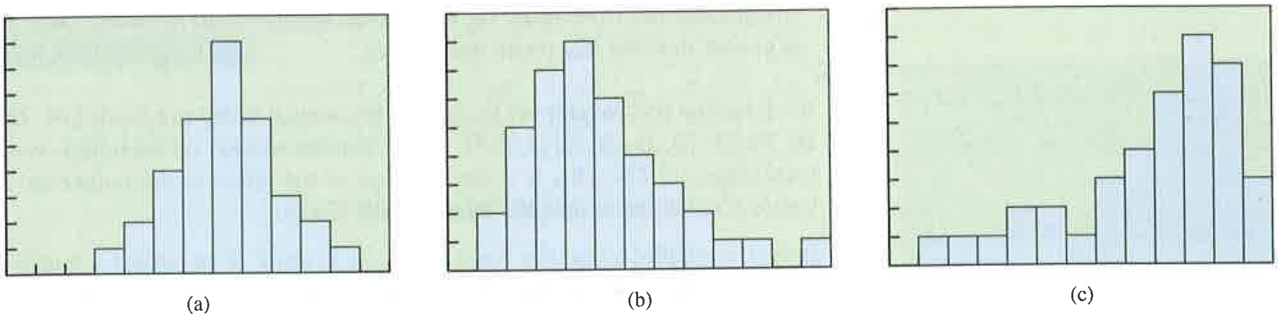


FIGURE 9.20 Which graph shows a data set in which the mean is less than the median? Greater than the median? Approximately equal to the median? (Exploration 1)

The distribution in Figure 9.20a is **symmetric** because it looks approximately the same when reflected about a vertical line through the median. The distribution in Figure 9.20b is **skewed right** because it has a longer "tail" to the right. The values in the tail will pull nonresistant measures (like the mean) to the *right*, leaving resistant measures (like the median) behind. The distribution in Figure 9.20c is **skewed left**, because nonresistant measures are pulled to the *left*.

You can also see symmetry and skewedness in stemplots, which have the same shapes vertically as histograms have horizontally.

Boxplots

A **boxplot** (sometimes called a **box-and-whisker plot**) is a graph that depicts the five-number summary of a data set. The plot consists of a central rectangle (box) that extends from the first quartile to the third quartile, with a vertical segment marking the median. Line segments (whiskers) extend at the ends of the box to the minimum and maximum values. For example, the five-number summary for the male life expectancies in South American nations (Example 7) was {59.0, 64.1, 68.75, 71.65, 72.6}. The boxplot for the data is shown in Figure 9.21. Notice that the box and the whisker extend further to the left of the median than to the right, suggesting a distribution that is skewed left. (The histogram obtained in Exercise 9 of Section 9.6 confirms that this is the case.)

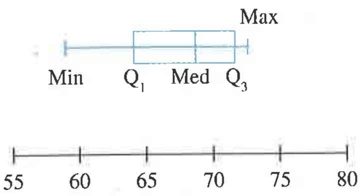


FIGURE 9.21 A boxplot for the five-number summary of the male life expectancies in Example 7. (The features on the box are labeled here for illustrative purposes; it is not necessary in general to label the min, the quartiles, or the max.)

EXAMPLE 8 Comparing boxplots

Draw boxplots for the male and female data in Example 7 and describe briefly the information displayed in the visualization.

SOLUTION The five-number summaries are:

Males: {59.0, 64.1, 68.75, 71.65, 72.6}

Females: {66.2, 70.25, 74.5, 77.7, 79.4}

The boxplots can be graphed simultaneously (Figure 9.22).

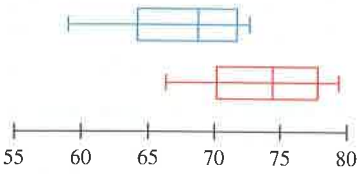
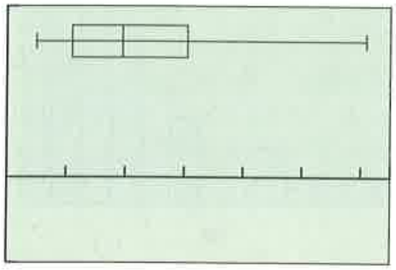


FIGURE 9.22 A single graph showing boxplots for male and female life expectancies in the nations of South America gives a good visualization of the differences in the two data sets. (Example 8)

From this graph we can see that the middle half of the female life expectancies are all greater than the median of the male life expectancies. The median life expectancy for the women among South American nations is greater than the maximum for the men. **Now try Exercise 31.**

If we look at the boxplot for Roger Maris's annual home run totals {14, 28, 16, 39, 61, 33, 23, 26, 8, 13, 9, 5}, we see that the whisker on the right is very long (Figure 9.23). This is a visualization of the effect of the outlier (61), which is much larger than the third quartile (30.5).



[0, 65] by [−5, 10]

FIGURE 9.23 A boxplot of Roger Maris's annual home run production (Table 9.11 on page 768). The outlier (61) results in a long whisker on the right because the maximum is much larger than Q_3.

In fact, a boxplot gives us a convenient way to think of an outlier: a number that makes one of the whiskers noticeably longer than the box. The usual rule of thumb for "noticeably longer" is 1.5 times as long. Since the length of the box is the *IQR*, that leads us to the following numerical check.

> A number in a data set can be considered an **outlier** if it is more than $1.5 \times IQR$ below the first quartile or above the third quartile.

EXAMPLE 9 Identifying an outlier

Is 61 an outlier in Roger Maris's home run data according to the $1.5 \times IQR$ criterion?

SOLUTION Maris's totals, in order: {5, 8, 9, 13, 14, 16, 23, 26, 28, 33, 39, 61}

His five-number summary: {5, 11, 19.5, 30.5, 61}

His *IQR*: $30.5 - 11 = 19.5$

So,

$$Q_3 + 1.5 \times IQR = 30.5 + 1.5 \times 19.5 = 59.75$$

Since $61 > 59.75$, the rule of thumb identifies it as an outlier.

Now try Exercise 39.

By their very nature, outliers can distort the overall picture we get of the data. For that reason, statisticians will frequently look for reasons to omit them from

DEFINING OUTLIERS

It must be pointed out that the rule of thumb given here for identifying outliers is not a universal definition. The only sure way to characterize an outlier is that it lies outside the expected range of the data, and that "expected range" can be a judgment call.

their statistical displays and calculations. (This can, of course, be risky. You want to omit a strange laboratory reading if you suspect equipment error, but you do not want to ignore a potential scientific discovery.) A **modified boxplot** is a compromise visualization that separates outliers as isolated points, extending the whiskers only to the farthest nonoutliers. Figure 9.24 shows a modified boxplot of Roger Maris's home run data, as compared to a regular boxplot.

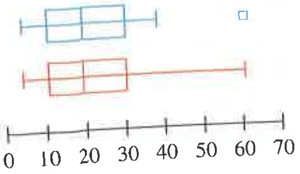


FIGURE 9.24 A modified boxplot and a regular boxplot of Roger Maris's annual home run totals.

Variance and Standard Deviation

You might be surprised that the five-number summary and its boxplot graph do not even make reference to the mean, which is a more familiar measure than a median or a quartile. This is because the mean, being a nonresistant measure, is less reliable in the presence of outliers or skewed data.

On the other hand, the mean is an excellent measure of center when outliers and skewedness are not present, which is quite often the case. Indeed, histograms of data from all kinds of real-world sources tend to look something like Figure 9.25, in which frequencies are higher close to the mean and lower as you move away from the mean in either direction. Statisticians call these *normal* distributions. (We will make that term more precise shortly.)

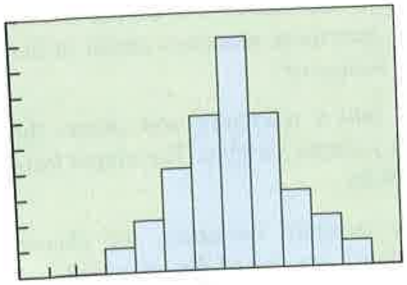


FIGURE 9.25 Histograms based on data gathered from real-world sources are often symmetric and higher in the middle, without outliers. Frequency distributions with graphs of this shape are called *normal*.

For normally distributed data, the mean is the preferred measure of center. There is also a measure of variability for normal data that is better than the *IQR*, called the *standard deviation*. Like the mean, the standard deviation is strongly affected by outliers and can be misleading if outliers are present.

Definition Standard Deviation

The **standard deviation** of the numbers $\{x_1, x_2, \ldots, x_n\}$ is

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

where $\bar{x}$ denotes the mean. The **variance** is σ^2, the square of the standard deviation.

If we define the "deviation" of a data value to be how much it differs from the mean, then the variance is just the mean of the squared deviations. The standard deviation is the square *root* of the *mean* of the *squared deviations*, which is why it is sometimes called the *root mean square deviation*. The symbol "σ" is a lowercase Greek letter sigma.

Calculating a standard deviation by hand can be tedious, but with modern calculators it is usually only necessary to enter the list of data and push a button. In fact, most calculators give you a choice of two standard deviations, one slightly larger than the other. The larger one (usually called s) is based on the formula

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

The difference is that the σ formula is for finding the true parameter, which means that it only applies when $\{x_1, x_2, \ldots, x_n\}$ is the whole **population**. If $\{x_1, x_2, \ldots, x_n\}$ is a *sample* from the population, then the s formula actually gives a better estimate of the parameter than the σ formula does. So use the larger standard deviation when your data come from a sample (which is almost always the case).

EXAMPLE 10 Finding standard deviation with a calculator

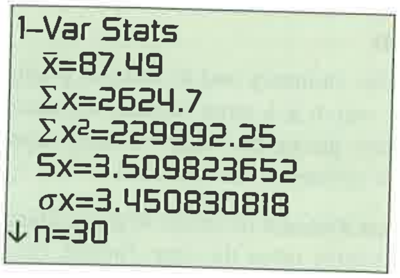
A researcher measured 30 newly hatched loon chicks and recorded their weights in grams as shown in Table 9.18.

TABLE 9.18 WEIGHTS IN GRAMS OF 30 LOON CHICKS

79.5	87.5	88.5	89.2	91.6	84.5	82.1	82.3	85.7	89.8
84.0	84.8	88.2	88.2	82.9	89.8	89.2	94.1	88.0	91.1
91.8	87.0	87.7	88.0	85.4	94.4	91.3	86.4	85.7	86.0

Based on the sample, estimate the mean and standard deviation for the weights of newly hatched loon chicks. Are these measures useful in this case, or should we use the five-number summary?

SOLUTION We enter the list of data into a calculator and choose the command that will produce statistics of a single variable. The output from one such calculator is shown in Figure 9.26.

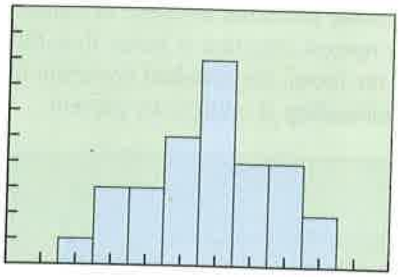


```
1-Var Stats
 x̄=87.49
 Σx=2624.7
 Σx²=229992.25
 Sx=3.509823652
 σx=3.450830818
↓n=30
```

FIGURE 9.26 Single-variable statistics in a typical calculator display. (Example 10)

The mean is $\bar{x} = 87.49$ grams. For standard deviation, we choose $Sx = 3.51$ grams because the calculations are based on a sample of loon chicks, not the entire population of loon chicks.

A histogram (Xscl = 2) in the window [77, 98] by [0, 10] shows that the distribution is normal (as we would expect from nature), containing no outliers or strong skewedness. Therefore, the mean and standard deviation are appropriate measures (Figure 9.27). Now try Exercise 35.



[77, 98] by [0, 10]

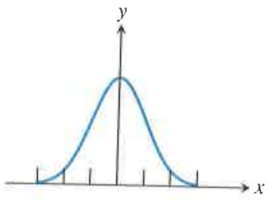
FIGURE 9.27 The weights of the loon chicks in Example 10 appear to be normal, with no outliers or strong skewedness. We conclude that the mean and standard deviation are appropriate measures of center and variability, respectively.

Normal Distributions

Although we use the word *normal* in many contexts to suggest typical behavior, in the context of statistics and data distributions it is really a technical term. If you graph the function

$$y = e^{-x^2/2}$$

in the window $[-3, 3]$ by $[0, 1]$, you will see what **normal** means mathematically (Figure 9.28).



(a)

FIGURE 9.28 The graph of $y = e^{-x^2/2}$. This is a Gaussian (or normal) curve.

The shape corresponds to the kind of distribution we have been calling "normal." In fact, this curve, called a **Gaussian curve** or **normal curve** is a *precise mathematical model* for normal behavior. That is where the mean and standard deviation come in.

The standard deviation of the curve in Figure 9.28 is 1. Using calculus, we can find that about 68% of the total area under this curve lies between −1 and 1 (Figure 9.29a). Since any normal distribution has this shape, *about 68% of the data in any normal distribution lie within 1 standard deviation of the mean.*

Similarly, we can find that about 95% of the total area under the Gaussian curve lies between −2 and 2 (Figure 9.29b), implying that *about 95% of the data in any normal distribution lie within 2 standard deviations of the mean.*

Similarly, we can find that about 99.7% (nearly all) of the total area under the Gaussian curve lies between −3 and 3 (Figure 9.29c), implying that *about 99.7% of the data in any normal distribution lie within 3 standard deviations of the mean.*

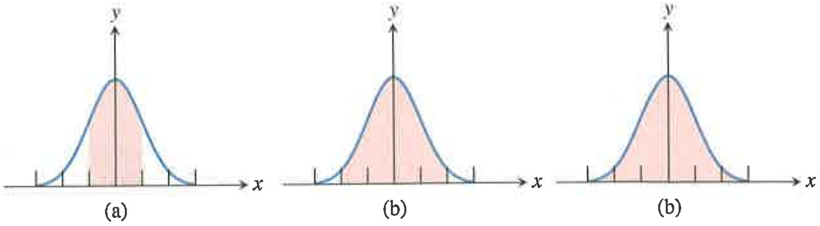


FIGURE 9.29 (a) About 68% of the area under a Gaussian curve lies within 1 unit of the mean. (b) About 95% of the area lies within 2 units of the mean. (c) About 99.7% of the area lies within 3 units of the mean. If we think of the units as standard deviations, this gives us a model for *any* normal distribution.

The 68–95–99.7 Rule

If the data for a population are normally distributed with mean μ and standard deviation σ, then

- approximately 68% of the data lie between $\mu - 1\sigma$ and $\mu + 1\sigma$;
- approximately 95% of the data lie between $\mu - 2\sigma$ and $\mu + 2\sigma$;
- approximately 99.7% of the data lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.

What makes this rule so useful is that normal distributions are common in a wide variety of statistical applications. We close the chapter with a simple application.

EXAMPLE 11 Using the 68–95–99.7 rule

Based on the research data presented in Example 10, would a loon chick weighing 95 grams be in the top 2.5% of all newly hatched loon chicks?

SOLUTION We assume that the weights of newly hatched loon chicks are normally distributed in the whole population. Since we do not know the mean and standard deviation for the whole population (the parameters μ and σ), we use $\bar{x} = 87.49$ and $Sx = 3.51$ as estimates.

Look at Figure 9.29b. The shaded region contains 95% of the area, so the two identical white regions at either end must each contain 2.5% of the area. That is, to be in the top 2.5%, a loon chick will have to weigh at least 2 standard deviations more than the mean:

$$\bar{x} + 2Sx = 87.49 + 2(3.51) = 94.51 \text{ grams.}$$

Since $95 > 94.51$, a 95-gram loon chick is indeed in the top 2.5%.

Now try Exercise 41.

If you study statistics more deeply someday you will learn that there is more going on in Example 11 than meets the eye. For starters, we need to know that the chicks are really a random sample of all loon chicks (not, for example, from the same geographical area). Also, we lose some accuracy by using a statistic to estimate the true standard deviation, and statisticians have ways of taking that into account.

This section has only offered a brief glimpse into how statisticians use mathematics. If you are interested in learning more, we urge you to find a good statistics textbook and pursue it!

QUICK REVIEW 9.7

(Prerequisite skill Section 9.4)

In Exercises 1–6, write the sum in expanded form.

1. $\sum_{i=1}^{7} x_i$

2. $\sum_{i=1}^{5} (x_i - \bar{x})$

3. $\frac{1}{7}\sum_{i=1}^{7} x_i$

4. $\frac{1}{5}\sum_{i=1}^{5} (x_i - \bar{x})$

5. $\frac{1}{5}\sum_{i=1}^{5} (x_i - \bar{x})^2$

6. $\sqrt{\frac{1}{5}\sum_{i=1}^{5} (x_i - \bar{x})^2}$

In Exercises 7–10, write the sum in sigma notation.

7. $x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_8 f_8$

8. $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_{10} - \bar{x})^2$

9. $\frac{1}{50}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_{50} - \bar{x})^2]$

10. $\sqrt{\frac{1}{7}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_7 - \bar{x})^2]}$

SECTION 9.7 EXERCISES

1. In each case, identify whether the number described is a *parameter* or a *statistic*.

(a) The average score on last week's quiz was 73.4.

(b) About 13% of the human population is left-handed.

(c) In a study of laboratory rats, 93% became aggressive when deprived of sleep.

2. In each case, identify whether the **average** described is a *mean*, *median*, or *mode*.

(a) The **average** Irish child with red hair also has freckles.

(b) The pitcher's earned run **average** is 2.35.

(c) The choir lined up with tall people in the back, short people in front, and people of **average** height in the middle.

In Exercises 3–6, find the mean of the data set.

3. {12, 23, 15, 48, 36}

4. {4, 8, 11, 6, 21, 7}

5. {32.4, 48.1, 85.3, 67.2, 72.4, 55.3}

6. {27.4, 3.1, 9.7, 32.3, 12.8, 39.4, 73.7}

In Exercises 7 and 8, find the mean population of the six states listed in Table 9.15 (page 769) for the indicated year.

7. 1900

8. 1990

In Exercises 9 and 10, find the mean of the indicated data.

9. The number of satellites (moons), from the data in Table 9.19.

TABLE 9.19 PLANETARY SATELLITES

Planet	Number of Satellites
Mercury	0
Venus	0
Earth	1
Mars	2
Jupiter	28
Saturn	30
Uranus	21
Neptune	8
Pluto	1

Source: World Almanac and Book of Facts, 2002.

10. The area of the continents, from the data in Table 9.20.

TABLE 9.20 SIZE OF CONTINENT

Continent	Area (km^2)
Africa	30,065,000
Antarctica	13,209,000
Asia	44,579,000
Australia/Oceania	7,687,000
Europe	9,938,000
North America	24,256,000
South America	17,819,000

Source: Worldatlas.com, 2002.

11. Find the median of the data in Table 9.19. (See Exercise 9.)

12. Find the median of the data in Table 9.20. (See Exercise 10.)

13. **Home Run Production** Determine the average annual home run production for Willie Mays and for Mickey Mantle for their career totals of 660 over 22 years and 536 over 18 years, respectively. Who had the greater production rate?

14. **Painting Houses** A painting crew in State College, Pennsylvania, painted 12 houses in 5 days, and a crew in College Station, Texas, painted 15 houses in 7 days. Determine the average number of houses per day each crew painted. Which crew had the greater rate?

15. **Skirt Production** The Hip-Hop House produced 1147 scooter skirts in 4 weeks, and What-Next Fashion produced 1516 scooter skirts in 4 weeks. Which company had the greater production rate?

16. **Per Capita Income** Per capita income (PCI) is an average found by dividing a nation's gross national product (GNP) by its population. India has 882,575,000 people and a GNP of 311 billion dollars, and Mexico has 87,715,000 people and a GNP of 218 billion dollars. Determine the PCI for India and for Mexico. Which nation has the greater income per person?

17. Find the median and mode of the numbers in Table 9.18 on page 780 (weights in grams of newly-hatched loon chicks).

18. In 1998, a total of 116,517 students took the Advanced Placement examination in Calculus AB. From the frequency table below, find the mean score in Calculus AB in 1998. *(Source: The College Board.)*

AP Examination Grade	Number of Students
5	18,522
4	27,102
3	31,286
2	20,732
1	18,875

19. In 1998, a total of 26,784 students took the Advanced Placement examination in Calculus BC. From the frequency table below, find the mean score in Calculus BC in 1998. *(Source: The College Board.)*

AP Examination Grade	Number of Students
5	9879
4	5119
3	6143
2	2616
1	3027

In Exercises 20 and 21 use the data in Table 9.16 (page 770).

(a) Find the average (mean) of the indicated temperatures for Beijing.

(b) Find the weighted average using the number of days in the month as the weight. (Assume no leap year.)

(c) Compare your results in (a) and (b). Do the weights have an effect on the average? Why or why not? Which average is the better indicator for these temperatures?

20. The monthly high temperatures

21. The monthly low temperatures

In Exercises 22–25, determine the five-number summary, the range, and the interquartile range for the sets of data specified. Identify any outliers.

22. The annual home run production data for Mark McGwire and Barry Bonds in Table 9.6 (page 763).

23. The annual home run production data for Willie Mays and Mickey Mantle in Table 9.13 (page 768).

24. The following average annual wind speeds at 44 climatic data centers around the United States:

9.0, 6.9, 9.1, 9.2, 10.2, 12.5, 12.0, 11.2, 12.9, 10.3, 10.6, 10.9, 8.7, 10.3, 11.0, 7.7, 11.4, 7.9, 9.6, 8.0, 10.7, 9.3, 7.9, 6.2, 8.3, 8.9, 9.3, 11.6, 10.6, 9.0, 8.2, 9.4, 10.6, 9.5, 6.3, 9.1, 7.9, 9.7, 8.8, 6.9, 8.7, 9.0, 8.9, 9.3

25. The following salaries for employees in one department of the Garcia Brothers Company (in thousands of dollars):

33.5, 35.3, 33.8, 29.3, 36.7, 32.8, 31.7, 37.3, 33.5, 28.2, 34.8, 33.5, 29.7, 38.5, 32.7, 34.8, 34.2, 31.6, 35.4

In Exercises 26–28, make **(a)** a boxplot and **(b)** a modified boxplot for the data.

26. The annual home run production data for Mark McGwire in Table 9.6 (page 763)

27. The annual home run production data for Barry Bonds in Table 9.6 (page 763)

28. The annual CD shipments in Table 9.9 (page 766)

In Exercises 29 and 30, refer to the wind speed data analyzed in Exercise 24.

29. Some wind turbine generators, to be efficient generators of power, require average wind speeds of at least 10.5 mph. Approximately what fraction of the climatic centers are suited for these wind turbine generators?

30. If technology improves the efficiency of the wind turbines so that they are efficient in winds that average at least 7.5 mph, approximately what fraction of the climatic centers are suited for these improved wind generators?

In Exercises 31 and 32, use simultaneous boxplots of the annual home run production data for Willie Mays and Mickey Mantle in Table 9.13 (page 768) to answer the question.

31. **(a)** Which data set has the greater range?

(b) Which data set has the greater interquartile range?

32. **Writing to Learn** Write a paragraph explaining the difference in home run production between Willie Mays and Mickey Mantle.

In Exercises 33–38, find the standard deviation and variance of the data set. (Since the data set is the population under consideration, use σ in each case, rather than s.)

33. $\{23, 45, 29, 34, 39, 41, 19, 22\}$

34. $\{28, 84, 67, 71, 92, 37, 45, 32, 74, 96\}$

35. The CD shipment data in Table 9.9 (page 766)

36. The cassette shipment data in Table 9.10 (page 767)

37. The data in Exercise 24

38. The data in Exercise 25

39. Mark McGwire's annual home run production (Table 9.6, page 763) ranges from a low of 3 to a high of 70. Is either of these numbers an outlier by the 1.5 × IQR criterion?

40. **Writing to Learn** Is it possible for the standard deviation of a set to be negative? To be zero? Explain your answer in both cases.

41. **SAT Scores** In 1995, scores on the Scholastic Aptitude Tests were re-scaled to their original mean (500) with an approximate standard deviation of 100. SAT scores in the general population have a normal distribution.

(a) Approximately what percentage of the 1995 scores fell between 400 and 600?

(b) Approximately what percentage of the 1995 scores fell below 300?

(c) By 2001 the national average on the SAT Math section had risen to 514. Is this number a parameter or a statistic?

42. **ACT Scores** In 2001, the national mean ACT Math score was 20.7, with an approximate standard deviation of 6. ACT scores in the general population have a normal distribution.

(a) Approximately what percentage of the 2001 scores were higher than 26.7?

(b) Approximately what ACT Math score would one need to make in 2001 to be ranked among the top 2.5% of all who took the test?

(c) **Writing to Learn** Mean ACT scores are published state by state. If we add up the 50 state means and divide by 50, will the result be a good estimate for the national mean score? Explain your answer.

Standardized Test Questions

43. **True or False** The median is strongly affected by outliers. Justify your answer.

44. **True or False** The length of the box in a boxplot is the interquartile range. Justify your answer.

You may use a graphing calculator when solving Exercises 45–48.

45. The plot of a normal distribution will be

(a) symmetric.

(b) skewed left.

(c) skewed right.

(d) lower in the middle than at the ends.

(e) of no predictable shape.

46. A frequency table for a set of 25 quiz grades is given below. What is the mean of the data?

Quiz Grade	10	9	8	7	6	5	4	3	2	1
Number of Students	3	3	5	6	4	3	1	0	0	0

(a) 7.00

(b) 7.28

(c) 7.35

(d) 7.60

(e) 7.86

47. Professor Mitchell grades the exams of his 30 students and finds that the scores have a mean of 81.3 and a median of 80.5. He later determines that the top student deserves 9 extra points for a misgraded proof. After the error is corrected, the scores have

(a) a mean of 81.3 and a median of 80.5.

(b) a mean of 81.6 and a median of 80.5.

(c) a mean of 81.6 and a median of 80.8.

(d) a mean of 90.3 and a median of 80.5.

(e) a mean of 90.3 and a median of 89.5.

48. If the calorie contents of robin eggs are normally distributed with a mean of 25 and a standard deviation of 1.2, then 95% of all robin eggs will have a calorie content in the interval

(a) [21.4, 28.6].

(b) [22, 28].

(c) [22.6, 27.4].

(d) [23, 27].

(e) [23.8, 26.2].

Explorations

49. **Group Activity** List a set of data for which the inequality holds.

(a) Mode $<$ median $<$ mean

(b) Median $<$ mean $<$ mode

(c) Mean $<$ mode $<$ median

50. **Group Activity** List a set of data for which the equation or inequality holds.

(a) Standard deviation $<$ interquartile range

(b) Interquartile range $<$ standard deviation

(c) Range $=$ interquartile range

51. Is it possible for the standard deviation of a data set to be greater than the range? Explain.

52. Why can we find the mode of categorical data but not the mean or median?

53. Draw a boxplot for which the inequality holds.

  (a) Median < mean

  (b) 2 × interquartile range < range

  (c) Range < 2 × interquartile range

54. Construct a set of data with median 5, mode 6, and mean 7.

Extending the Ideas

Weighting Data by Population The average life expectancies for males and females in South American nations were given in Table 9.12 (page 768). To find an overall average life expectancy for males or females in all of these nations, we would need to weight the national data according to the various national populations. Table 9.21 is an extension of Table 9.12, showing the populations (in millions). Assume that males and females appear in roughly equal numbers in every nation.

TABLE 9.21 LIFE EXPECTANCY BY GENDER AND POPULATION (IN MILLIONS) FOR THE NATIONS OF SOUTH AMERICA

Nation	Male	Female	Population
Argentina	71.9	78.8	37.4
Bolivia	61.5	66.7	8.3
Brazil	59.0	67.7	174.5
Chile	72.6	79.4	15.3
Colombia	66.7	74.6	40.3
Ecuador	68.5	74.3	13.2
Guyana	60.5	66.2	0.7
Paraguay	71.4	76.5	5.7
Peru	67.9	72.8	27.5
Suriname	69.0	74.4	0.4
Uruguay	72.1	79.0	3.4
Venezuela	70.3	76.6	23.9

Source: The World Almanac and Book of Facts, 2002.

In Exercises 55 and 56, use the data in Table 9.21 to find the mean life expectancy for each group.

55. Women living in South American nations

56. Men living in South American nations

57. **Quality Control** A plant manufactures ball bearings to the purchaser's specifications, rejecting any output with a diameter that deviates more than 0.1 mm from the specified value. If the ball bearings are produced with the specified mean and a standard deviation of 0.05 mm, what percentage of the output will be rejected?

58. **Quality Control** A machine fills 12-ounce cola cans with a mean of 12.08 ounces of cola and a standard deviation of 0.04 ounces. Approximately what percentage of the cans will actually contain less than the advertised 12 ounces of cola?

Math at Work

In the course of getting my bachelor's degree in communications, I took a class in math communications research, and I enjoyed it immensely. After I graduated from college, I interviewed in print journalism, but I wasn't a good editor. Since I really enjoyed research and analysis, I got started in research management.

What I like best about my job is that I get to act as a detective, using math to find out about people. Basically, my job is to bring the voice of the customers into a product development plan, so we can figure out how to make a product for them. There are two ways to get to know your customers: mathematics and psychology. Using psychology, the people developing the product have a tendency to use their gut feelings, which are sometimes entirely incorrect, to surmise what the customers want. That's why we use mathematics.

We use a set of questions, called a "Customer Satisfaction Survey" to find out what values are important to the customers. Then we can analyze the raw information and find out such things as the differences between frequent and infrequent users, and how males and females differ in their uses of the product. In this way, I learn all about people, and we, as a company, know how to tailor a product to our customers.

VERONICA GUERRERRO

CHAPTER 9 Key Ideas

PROPERTIES, THEOREMS, AND FORMULAS

Multiplication Principle of Counting (p. 704)
Permutation Counting Formula (p. 706)
Combination Counting Formula (p. 707)
Formula for Counting Subsets of an n-Set (p. 709)
Recursion Formula for Pascal's Triangle (p. 715)
The Binomial Theorem (p. 716)
Basic Factorial Identities (p. 716)
Probability of an Event (Equally Likely Outcomes) (p. 720)
Probability of an Event (Outcomes Not Equally Likely) (p. 721)
Multiplication Principle of Probability (p. 723)
Conditional Probability Formula (p. 726)
Sum of a Finite Arithmetic Sequence Theorem (p. 740)
Sum of a Finite Geometric Sequence Theorem (p. 742)
Sum of an Infinite Geometric Series Theorem (p. 745)
Principle of Mathematical Induction (p. 752)
The 68-95-99.7 Rule (p. 781)

PROCEDURES

Strategy for Determining Probabilities (p. 722)

CHAPTER 9 Review Exercises

The collection of exercises marked in red could be used as a chapter test.

In Exercises 1–6, evaluate the expression by hand, then check your result with a calculator.

1. $\binom{12}{5}$

2. $\binom{789}{787}$

3. ${}_{18}C_{12}$

4. ${}_{35}C_{28}$

5. ${}_{12}P_7$

6. ${}_{15}P_8$

7. **Code Words** How many five-character code words are there if the first character is always a letter and the other characters are letters and/or digits?

8. **Scheduling Trips** A travel agent is trying to schedule a client's trip from city A to city B. There are three direct flights, three flights from A to a connecting city C, and four flights from this connecting city C to city B. How many trips are possible?

9. **License Plates** How many license plates begin with two letters followed by four digits or begin with three digits followed by three letters? Assume that no letters or digits are repeated.

10. **Forming Committees** A club has 45 members, and its membership committee has three members. How many different membership committees are possible?

11. **Bridge Hands** How many 13-card bridge hands include the ace, king, and queen of spades?

12. **Bridge Hands** How many 13-card bridge hands include all four aces and exactly one king?

13. **Coin Toss** Suppose that a coin is tossed five times. How many different outcomes include at least two heads?

14. **Forming Committees** A certain small business has 35 employees, 21 women and 14 men. How many different employee representative committees are there if the committee must consist of two women and two men?

15. **Code Words** How many code words of any length can be spelled out using game tiles of five different letters (including single-letter code words)?

16. **A Pocket of Coins** Sean tells Moira that he has less than 50 cents in American coins in his pocket and no two coins of the same denomination. How many possible total amounts could be in Sean's pocket?

17. **Permutations** Find the number of distinguishable permutations that can be made from the letters in

(a) GERMANY

(b) PRESBYTERIANS

In each case, can you find a permutation that spells the first and last name of a well-known female entertainer?

18. **Permutations** Find the number of distinguishable permutations that can be made from the letters in

(a) FLORIDA

(b) TALLAHASSEE

In Exercises 19–24, expand each expression.

19. $(2x + y)^5$

20. $(4a - 3b)^7$

21. $(3x^2 + y^3)^5$

22. $\left(1 + \frac{1}{x}\right)^6$

23. $(2a^3 - b^2)^9$

24. $(x^{-2} + y^{-1})^4$

25. Find the coefficient of x^8 in the expansion of $(x - 2)^{11}$.

26. Find the coefficient of x^2y^6 in the expansion of $(2x + y)^8$.

In Exercises 27–30, list the elements of the sample space.

27. **Spinners** A game spinner on a circular region divided into 6 equal sectors numbered 1–6 is spun.

28. **Rolling Dice** A red die and a green die are rolled.

29. **Code Words** A two-digit code is selected from the digits $\{1, 3, 6\}$, where no digits are to be repeated.

30. **Production Line** A product is inspected as it comes off the production line and is classified as either defective or nondefective.

In Exercises 31–34, a penny, a nickel, and a dime are tossed.

31. List all possible outcomes.

32. List all outcomes in the event "two heads or two tails."

33. List all outcomes in the complement of the event in Exercise 32 (i.e., the event "neither two heads nor two tails.")

34. Find the probability of tossing at least one head.

35. **Coin Toss** A fair coin is tossed six times. Find the probability of the event "HHTHTT."

36. **Coin Toss** A fair coin is tossed five times. Find the probability of obtaining two heads and three tails.

37. **Coin Toss** A fair coin is tossed four times. Find the probability of obtaining one head and three tails.

38. **Assembly Line** In a random check on an assembly line, the probability of finding a defective item is 0.003. Find the probability of a nondefective item occurring 10 times in a row.

39. **Success or Failure** An experiment has only two possible outcomes—success (S) or failure (F)—and repetitions of the experiment are independent events. If $P(S) = 0.5$, find the probability of obtaining three successes and one failure in four repetitions.

40. **Success or Failure** For the experiment in Exercise 39, explain why the probability of one success and three failures is equal to the probability of three successes and one failure.

In Exercises 41–44, an experiment has only two possible outcomes—success (S) or failure (F)—and repetitions are independent events. The probability of success is 0.4.

41. Find the probability of SF on two repetitions.

42. Find the probability of SFS on three repetitions.

43. Find the probability of at least one success on two repetitions.

44. Explain why the probability of one success and three failures is not equal to the probability of three successes and one failure.

45. Mixed Nuts Two cans of mixed nuts of different brands are open on a table. Brand A consists of 30% cashews, while brand B consists of 40% cashews. A can is chosen at random, and a nut is chosen at random from the can. Find the probability that the nut is

(a) from the brand A can.

(b) a brand A cashew.

(c) a cashew.

(d) from the brand A can, given that it is a cashew.

46. Horse Racing If the track is wet, Mudder Earth has a 70% chance of winning the fifth race at Keeneland. If the track is dry, she only has a 40% chance of winning. Weather forecasts predict an 80% chance that the track will be wet. Find the probability that

(a) the track is wet and Mudder Earth wins.

(b) the track is dry and Mudder Earth wins.

(c) Mudder Earth wins.

(d) (in retrospect) the track was wet, given that Mudder Earth won.

In Exercises 47 and 48, find the first 6 terms and the 40th term of the sequence.

47. $a_n = \dfrac{n^2 - 1}{n + 1}$ **48.** $b_k = \dfrac{(-2)^k}{k + 1}$

In Exercises 49–54, find the first 6 terms and the 12th term of the sequence.

49. $a_1 = -1$ and $a_n = a_{n-1} + 3$, for $n \geq 2$

50. $b_1 = 5$ and $b_k = 2b_{k-1}$, for $k \geq 2$

51. Arithmetic sequence, with $a_1 = -5$ and $d = 1.5$

52. Geometric sequence, with $a_1 = 3$ and $r = 1/3$

53. $v_1 = -3$, $v_2 = 1$, and $v_k = v_{k-2} + v_{k-1}$, for $k \geq 3$

54. $w_1 = -3$, $w_2 = 2$, and $w_k = w_{k-2} + w_{k-1}$, for $k \geq 3$

In Exercises 55–62, the sequences are arithmetic or geometric. Find an explicit formula for the nth term. State the common difference or ratio.

55. 12, 9.5, 7, 4.5, . . . **56.** −5, −1, 3, 7, . . .

57. 10, 12, 14.4, 17.28, . . . **58.** $\frac{1}{8}, -\frac{1}{4}, \frac{1}{2}, -1, \ldots$

59. $a_1 = -11$ and $a_n = a_{n-1} + 4.5$ for $n \geq 2$

60. $b_1 = 7$ and $b_n = (1/4)\, b_{n-1}$ for $n \geq 2$

61. The fourth and ninth terms of a geometric sequence are −192 and 196,608, respectively.

62. The third and eighth terms of an arithmetic sequence are 14 and −3.5, respectively.

In Exercises 63–66, find the sum of the terms of the arithmetic sequence.

63. −11, −8, −5, −2, 1, 4, 7, 10

64. 13, 9, 5, 1, −3, −7, −11

65. 2.5, −0.5, −3.5, . . . , −75.5

66. −5, −3, −1, 1, . . . , 55

In Exercises 67–70, find the sum of the terms of the geometric sequence.

67. $4, -2, 1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}$

68. $-3, -1, -\frac{1}{3}, -\frac{1}{9}, -\frac{1}{27}$

69. 2, 6, 18, . . . , 39,366

70. 1, −2, 4, −8, . . . , −8192

In Exercises 71 and 72, find the sum of the first 10 terms of the arithmetic or geometric sequence.

71. 2187, 729, 243, . . . **72.** 94, 91, 88, . . .

In Exercises 73 and 74, graph the sequence.

73. $a_n = 1 + \dfrac{(-1)^n}{n}$ **74.** $a_n = 2n^2 - 1$

75. Annuity Mr. Andalib pays $150 at the end of each month into an account that pays 8% interest compounded monthly. At the end of 10 years, the balance in the account, in dollars, is

$$150\left(1 + \frac{0.08}{12}\right)^0 + 150\left(1 + \frac{0.08}{12}\right)^1 + \cdots + 150\left(1 + \frac{0.08}{12}\right)^{119}.$$

Use the formula for the sum of a finite geometric series to find the balance.

76. Annuity What is the minimum monthly payment at month's end that must be made in an account that pays 8% interest compounded monthly if the balance at the end of 10 years is to be at least $30,000?

In Exercises 77–82, determine whether the geometric series converges. If it does, find its sum.

77. $\sum_{j=1}^{\infty} 2\left(\frac{3}{4}\right)^j$

78. $\sum_{k=1}^{\infty} 2\left(-\frac{1}{3}\right)^k$

79. $\sum_{j=1}^{\infty} 4\left(-\frac{4}{3}\right)^j$

80. $\sum_{k=1}^{\infty} 5\left(\frac{6}{5}\right)^k$

81. $\sum_{k=1}^{\infty} 3(0.5)^k$

82. $\sum_{k=1}^{\infty} (1.2)^k$

In Exercises 83–86, write the sum in sigma notation.

83. $-8 - 3 + 2 + \cdots + 92$

84. $4 - 8 + 16 - 32 + \cdots - 2048$

85. $1^2 + 3^2 + 5^2 + \cdots$

86. $1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots$

In Exercises 87–90, use summation formulas to evaluate the expression.

87. $\sum_{k=1}^{n} (3k + 1)$

88. $\sum_{k=1}^{n} 3k^2$

89. $\sum_{k=1}^{25} (k^2 - 3k + 4)$

90. $\sum_{k=1}^{175} (3k^2 - 5k + 1)$

In Exercises 91–94, use mathematical induction to prove that the statement is true for all positive integers *n*.

91. $1 + 3 + 6 + \cdots + \frac{n(n+1)}{2} = \frac{n(n+1)(n+2)}{6}$

92. $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \frac{n(n+1)(n+2)}{3}$

93. $2^{n-1} \leq n!$

94. $n^3 + 2n$ is divisible by 3.

In Exercises 95–98, construct **(a)** a stemplot, **(b)** a frequency table, and **(c)** a histogram for the indicated data.

95. Real Estate Prices Use intervals of \$10,000. The median sales prices (in units of \$10,000) for homes in 30 randomly selected metropolitan areas in 2001 were as follows:

10.7, 11.4, 12.7, 11.5, 14.6, 13.6, 9.2, 21.9, 16.1, 12.2, 13.5, 12.6, 12.0, 14.7, 23.4, 12.4, 17.0, 11.7, 11.5, 10.6, 14.1, 15.4, 15.8, 17.6, 14.7, 11.7, 12.7, 9.1, 16.4, 14.8

(Source: *National Association of Realtors,* as reported in *The World Almanac and Book of Facts, 2002.*)

96. Popular Web Sites Use intervals of 5 million. The number of visitors (in units of 10 million) to the top 25 Web sites in 2001 (as measured May 1–31) were as follows:

7.1, 6.1, 5.8, 3.3, 2.9, 2.8, 2.6, 2.0, 2.0, 2.0, 1.9, 1.9, 1.9, 1.5, 1.4, 1.4, 1.3, 1.3, 1.2, 1.2, 1.1, 1.1, 1.1, 1.1, 1.1

(Source: *Media Matrix,* as reported in *The New York Times Almanac, 2002.*)

97. Beatles Songs Use intervals of length 10. The lengths (in seconds) of 24 randomly selected Beatles songs that appeared on singles are as follows, in order of release date:

143, 120, 120, 139, 124, 144, 131, 132, 148, 163, 140, 177, 136, 124, 179, 131, 180, 137, 156, 202, 191, 197, 230, 190

(Source: Personal collection.)

98. Passing Yardage In 1995, Warren Moon of the Minnesota Vikings became the first pro quarterback to pass for 60,000 total yards. Use intervals of 1000 yards for Moon's regular season passing yards given in Table 9.25.

TABLE 9.22 REGULAR SEASON PASSING YARDAGE STATISTICS FOR WARREN MOON

Year	Yards	Year	Yards
1978	1112	1987	2806
1979	2382	1988	2327
1980	3127	1989	3631
1981	3959	1990	4689
1982	5000	1991	4690
1983	5648	1992	2521
1984	3338	1993	3485
1985	2709	1994	4264
1986	3489	1995	4228

Source: The Minnesota Vikings, as reported by Julie Stacey in USA Today on September 25, 1995 and www.nfl.com

In Exercises 99–102, find the five-number summary, the range, the interquartile range, the standard deviation, and the variance σ for the specified data (use σ and σ^2). Identify any outliers.

99. The data in Exercise 95

100. The data in Exercise 96

101. The data in Exercise 97

102. The data in Exercise 98

In Exercises 103–106, construct **(a)** a boxplot and **(b)** a modified boxplot for the specified data.

103. The data in Exercise 95

104. The data in Exercise 96

105. The data in Exercise 97

106. The data in Exercise 98

107. Make a back-to-back stemplot of the data in Exercise 97, showing the earlier 12 songs in one plot and the later 12 songs in the other. Write a sentence interpreting the stemplot.

108. Make simultaneous boxplots of the data in Exercise 97, showing the earlier 12 songs in one boxplot and the later 12 songs in the other.

(a) Which set of data has the greater range?

(b) Which set of data has the greater interquartile range?

109. Time Plots Make a time plot for the data in Exercise 97, assuming equal time intervals between songs. Interpret the trend revealed in the time plot.

110. Damped Time Plots Statisticians sometimes use a technique called *damping* to smooth out random fluctuations in a time plot. Find the mean of the first four numbers in Exercise 97, the mean of the next four numbers, and so on. Then graph the six means as a function of time. Is there a clear trend?

111. Find row 9 of Pascal's triangle.

112. Show algebraically that

$$_nP_k \times {}_{n-k}P_j = {}_nP_{k+j}.$$

113. Baseball Bats Suppose that the probability of producing a defective baseball bat is 0.02. Four bats are selected at random. What is the probability that the lot of four bats contains the following?

(a) No defective bats

(b) One defective bat

114. Light Bulbs Suppose that the probability of producing a defective light bulb is 0.0004. Ten light bulbs are selected at random. What is the probability that the lot of 10 contains the following?

(a) No defective light bulbs

(b) Two defective light bulbs

4, 7, 10 — — $(3K+1)$

$$\sum_{1}^{25} (K+4)(K-1)$$

CHAPTER 9 Project

Analyzing Height Data

The set of data below was gathered from a class of 30 precalculus students. Use this set of data or collect the data for your own class and use it for analysis.

Heights of students in inches				
66	69	72	64	68
70	71	66	65	63
72	59	64	63	66
68	63	64	71	71
69	62	61	67	69
64	73	75	61	70

1. Create a stem-and-leaf plot of the data using split stems. From this data, what is the approximate average height of a student in the class?

2. Create a frequency table for the data using an interval of 2. What information does this give?

3. Create a histogram for the data using an interval of 2. What conclusions can you draw from this representation of the data? Can you estimate the average height for males and the average height for females?

4. Compute the mean, median, and mode for the data set. Discuss whether each is a good measure of the average height of a student in the class. Is each a good predictor for average height of students in other precalculus classes?

5. What can you say about the data if the mean and median values are close?

6. Find the five-number summary for the class heights.

7. Create a boxplot and explain what information it gives about the data set.

8. A new student is now added to the class. He is a 7'2" star basketball player. Add his height to the data set. Recalculate the mean, median, and five-number summary. Create a new boxplot and use your calculator to plot it underneath the boxplot for the original class. How does this new student affect the statistics?

9. Explain why this new student would be considered an outlier and the importance of identifying outliers when calculating statistics and making predictions from them.

10. Suppose now that three additional basketball players transferred into the class. They are 7'0", 6'11" and 6'10". Recalculate the statistics from number 9 and discuss the implications of using these statistics to make predictions for other precalculus classes.