CHAPTER 4

PROBABILITY AND STATISTICS

- 4.1 Elementary and Conditional Probability
- 4.2 Odds and Expected Value
- 4.3 STATISTICAL GRAPHS
- 4.4 ANALYZING DATA
- 4.5 Polls and the Margin of Error

In this chapter, you will be introduced to probability and statistics. Probability will be presented from the point of view of its intuitive concepts. You will learn how these intuitive ideas can be connected to mathematical rules through the use of both sets and common sense. In the material on statistics, you will see how data can be organized, displayed, and analyzed. Our focus in this chapter will be on performing the basic statistical calculations and understanding their meaning.

Research Projects

- 1. What does a standard deck of playing cards consist of? What are the origins of playing cards? Give the origins and history of at least three card games, such as poker, blackjack, whist, bridge, pinochle, and euchre.
- 2. Research the profitability of gambling casinos and the effects that such casinos have on the economy of their local communities.
- 3. What is the meaning of the expression "lie with statistics"? Give some examples of this and explain how each of the examples is a "lie."

Math Projects

- 1. Examine one of the state-run lotteries. How is the lottery played? What are the odds or probabilities of winning in the lottery? What is the expected value for this lottery?
- 2. The games of chance in gambling casinos make use of probability. Write a short paper on one of the games following games: Keno, Roulette, Craps. Include a history of the game, a description of how the game is played, and an analysis of the probabilities, odds and expected values involved in the game.
- 3. You have been hired as a consultant by a strawberry farmer. After being awarded the contract, you receive the following letter.

To: Recently Hired Statistician

From: Morgan Hill Berry Growers

Please advise us on which company to use as our strawberry distributor. Four highly recommended distributors have provided us with statistical data on the weekly prices for one load of strawberries per week for a ten-week period last year. Prices fluctuate according to availability, and we would like to use the company with the lowest overall price and the least amount of fluctuation. We would like your written report showing your results and a detailed recommendation as to which company we should choose. Thank you.

			AlwaysRipe	BerryDelicious
Week	Fresh Picked, Inc.	FastNFresh	Fruit Company	Deliveries
1	\$355	\$350	\$350	\$360
2	\$350	\$345	\$350	\$300
3	\$310	\$295	\$320	\$320
4	\$330	\$325	\$320	\$320
5	\$340	\$315	\$330	\$290
6	\$290	\$290	\$300	\$305
7	\$305	\$305	\$310	\$290
8	\$315	\$300	\$315	\$310
9	\$325	\$315	\$345	\$340
10	\$350	\$340	\$290	\$345

What is your recommendation to Morgan Hill Berry Growers?

Section 4.1 Elementary and Conditional Probability

In everyday conversations, we frequently ask questions such as "What's my chance of getting an A in this class?" "What's the probability of my winning the drawing for the trip to Hawaii?" "What is the chance that it will rain today?" What we are looking for is some kind of a measure of the chance that the event will occur. I might say that there is a 90% chance that I will get an A in the class, or the probability of my winning the drawing for the Hawaiian vacation is 1 in 20,000, or there is a 50% chance of rain today. The numbers measure the likelihood that the event will occur.

For example, the probability of flipping a coin and having it land with the head side up is $\frac{1}{2}$. Since the coin could land on either heads or tails, the head outcome is one of the two possible outcomes. In general, the **probability** of an event is found by dividing the number of ways that an event can occur (number of desired outcomes) by the total number of possible outcomes. If we will let P(*E*) represent the probability of an event *E*, the basic formula for determining the probability is

$$P(E) = \frac{\text{number of ways an event can occur (desired)}}{\text{total number of possible outcomes (total)}} \text{ where } 0 \le P(E) \le 1$$

Example 1

Find the probability of drawing an ace from a standard deck of 52 cards.

Solution: In this case, the event is drawing an ace. Because there are four aces in the standard deck, there are four ways to draw an ace out of a total of 52 possibilities. $P(Ace) = \frac{4}{52} = \frac{1}{13}$

The probability of an event can be represented as a fraction, decimal, or percent. Each form gives a numerical way to analyze the chance that the event occurs. Thus, the probability of drawing an ace in Example 1 can be written $P(Ace) = \frac{1}{13} \approx 0.077 \approx 7.7\%$

Example 2

A question on a multiple-choice test has five answers. What is the probability that you guess the correct answer to the question?

Solution: There is only one correct (desired) answer out of the five possible (total) answers. Therefore, P(correct answer) = $\frac{1}{5}$

Example 3

An American Roulette wheel has thirty-eight compartments around its circumference. Thirty six compartments are numbered from 1 to 36 with half of them colored red and half colored black. The remaining two compartments, numbered 0 and 00, are colored green. A ball is spun and lands in one of the compartments.

- a) What is the probability that the ball lands on the number 27?
- b) What is the probability that the ball does not land on the number 27?
- c) What is the probability that the ball on the tenth spin lands on the number 27?

- d) What is the probability that the ball lands on an odd number?
- e) What is the probability that the ball lands on a green colored compartment?
- **Solutions**: For each part of this problem we determine the number of desired outcomes and the number of total possible outcomes. The probability is then calculated by dividing these results.
 - a) There is one desired outcome out of the 38 total outcomes. Therefore, $P(27) = \frac{1}{38}$
 - b) Since there are 37 ways for the ball to not land on the number 27, P(not 27) = $\frac{37}{32}$
 - c) The probability for the ball landing on the number 27 for the tenth or any other spin is the same. Therefore, $P(27) = \frac{1}{38}$
 - d) There are 18 odd-numbered compartments on the wheel (desired outcomes) out of the 38 total outcomes. Therefore, $P(odd) = \frac{18}{38} = \frac{9}{19}$
 - e) There are 2 green compartments on the wheel (desired outcomes) out of the 38 total outcomes. Therefore, P(green) = $\frac{2}{38} = \frac{1}{19}$

If we know the probability of an event, we can use this number to help predict the number of occurences of an event.

Example 4

Suppose that the probability of a student wearing blue jeans to school is 31%. In a class of 40 students, approximately how many students will be wearing blue jeans?

Solution: This is a type of percentage problem you may have encountered previously. It is asking for 31% of the 40 students so we can expected $0.31 \times 40 \approx 12$ students to be wearing blue jeans.

Two Restrictions on Probability

There are two restrictions on the value of a probability. First, in the formula for probability, the number on the bottom of the fraction must be greater than or equal to the number on the top of the fraction. Therefore, P(E) must be less than or equal to 1.

The second restriction is that since the number of ways in which an event occurs cannot be negative, P(E) must be greater than or equal to zero. Thus, we have $0 \le P(E) \le 1$.

The closer the probability is to zero, the less of a chance there is that the event will occur. The closer the probability is to 1, the more of a chance there is that the event will occur. If it is impossible for an event to occur, the probability of the event is 0. If an event is certain to happen, the probability of the event is 1. The next example demonstrates these concepts.

Example 5

Two standard dice are tossed.

- a) What is the probability that the sum of the two dice is 14?
- b) What is the probability that the sum of the two dice is less than 13?

Solution:

- a) Since the maximum number of dots on a die is six, it is impossible to get a sum of 14 with two standard dice. Therefore, P(sum of 14) = 0.
- b) Since the sum of two standard dice is always less than 13, the event is a certain event. The probability of an event that is certain is 1. Therefore, P(sum less than 13) = 1.

Conditional Probability

We continue the discussion of the probability by examining conditional probability. **Conditional probability** is the probability of an event occurring if some other condition has already occurred. The knowledge of this other condition often changes the probability. For example, the probability that you are more than six feet tall if your parents are both more than six feet tall is much higher than the probability that a person selected at random is more than six feet tall.

Example 6

Based on statistics published by the U.S. Department of Labor, the employment classifications of 1200 men and 1300 women are given in the following table.

	Number of	Number of	
Occupation	Males	Females	Total
Executive, administrative, management	180	185	365
Professional specialty	163	234	397
Technical, sales, administrative support	236	520	756
Service occupations	119	227	346
Precision production, crafts	223	27	250
Operators, fabricators, laborers	232	92	324
Farming, forestry, fishing	47	15	62
Totals	1200	1300	2500

- a) What is the probability of a person chosen at random being male if that person works in a service occupation?
- b) What is the probability of a person chosen at random working in a service occupation if that person is male?
- **Solution**: We will let *M* indicate males and *S* indicate service occupations. Notice that the wording of the two questions is reversed.
 - a) In part (a), it is known that the person works in a service occupation. Therefore, there are a total of 346 people of which, 119 are male. This gives

$$P(M \text{ if } S) = \frac{119}{346}$$

b) In part (b), it is known that the person is male. Since there is a total of 1200 male workers and 119 of them are in service occupations, $P(S \text{ if } M) = \frac{119}{1200}$.

Notice how the wording of the two parts is different and that this drastically affects the answer. It is important to read carefully when you are answering this type of question.

Example 7

A standard deck of cards contains 52 cards, including four queens. You are going to pick and keep two of the cards.

- a) What is the probability that the second card is a queen if the first card is a queen?
- b) What is the probability that the second card is a queen if the first card is not a queen?

Solution: If we let Q mean that we pick a queen and

- a) After you pick the first queen, there are still three queens in the deck out of a total of 51 cards. Therefore, P(2nd card Q if 1st card Q) = $\frac{3}{51} = \frac{1}{17}$.
- b) Since the first card you picked is not a queen, there are still four queens left in the deck out of a total of 51 cards. Therefore, $P(2^{nd} \text{ card } Q \text{ if } 1^{st} \text{ card not } Q) = \frac{4}{51}$.

§4.1 Explain \rightarrow Apply \rightarrow Explore Explain

- 1. What is probability?
- 2. Explain what P(E) = 0 means and describe an event that has a probability of zero.
- 3. Explain what P(E) = 1 means and describe an event that has a probability of 1.
- 4. What is the meaning of a political poll that says there is a 63% chance that the candidate will win reelection?
- 5. What is meant by conditional probability?
- 6. If A is the event that you like to eat apples and B is the event that you like to go bike riding, explain in words the meaning of P(B if you know A).

Apply

- 7. A card is drawn from a standard deck of 52 cards. (A standard deck contains 13 spades.)
 - a) What is the probability that the seven of spades is drawn?
 - b) What is the probability that a seven is drawn?
 - c) What is the probability that a face card (king, queen, or jack) is drawn?
 - d) What is the probability that a heart is drawn?
- 8. Two six-sided dice, with sides numbered 1 through 6, are rolled.
 - a) What is the probability that the sum of the two dice is 8?
 - b) What is the probability that the sum of the two dice is 1?
 - c) What is the probability that exactly one of the two dice shows a 3?
 - d) What is the probability that the sum of the two dice is 13?

- 9. In baseball, if a player has a batting average of 0.279, it means that for every 1,000 official at bats, the player gets a hit 279 times. Ty Cobb of the Detroit Tigers had a lifetime batting average of 0.367. How many hits would Ty expect in a season if he had 530 official at bats?
- 10. If there is a 75% chance that you will get up when the alarm rings, what is the probability that you will not get up when the alarm rings?
- 11. Martina believes that the probability of her passing her next test is 60%. What is the probability that Martina does not pass?

Explore

- 12. Suppose five standard decks of cards are combined.
 - a) What is the probability that the seven of spades is the first card drawn?
 - b) What is the probability that a seven is the first card drawn?
 - c) What is the probability that either an ace or an eight is the first card drawn?
 - d) Suppose that the first card drawn is an ace. What is the probability that the second card drawn is an ace?
- 13. The probability of gene mutation under certain conditions is 0.00006. What is the probability of a gene under these conditions not mutating?
- 14. A pinochle deck of cards consists of 48 cards: eight aces, eight kings, eight queens, eight jacks, eight tens, and eight nines. Suppose you are cutting cards from a pinochle deck.
 - a) What is the probability that you draw a king on the first card?
 - b) Suppose you draw a king on the first card and do not return the king to the deck before you draw a second card. What is the probability of drawing a king?
 - c) Suppose you draw another king on the second card and again you do not return the king to the deck before you draw a third time. What is the probability that you now draw a king?

There are 20 envelopes in a hat. Each envelope contains a piece of paper in the shape of a plane or a boat. On each slip of paper is written Hawaii or Mexico. You have been chosen to pick an envelope from the hat, and then you get to go on that type of vacation. The follow-ing table shows the number of each of the four types of vacations that are in the hat. Use this information to solve Problems 15 - 18.

	Hawaii	Mexico	Totals
Plane	7	8	15
Boat	2	3	5
Totals	9	11	20

- 15. What is the probability that you won a trip to Mexico if you won a plane flight?
- 16. What is the probability that you won a trip to Hawaii if you won a plane flight?
- 17. What is the probability that you won a cruise if you won a trip to Hawaii?
- 18. What is the probability that you won a plane trip if you won a trip to Mexico?

	Enrolled in					
Working	Fewer Than 6 Units	Between 6 and 13 Units	More than 13 Units			
Fewer Than 10 Hours/Week	28	72	85			
Between 10 and 20 Hours/Week	29	150	72			
More than 20 Hours/Week	150	82	62			

Use this table to answer the following questions.

- 19. What is the probability that if a student is working fewer than ten hours per week, the student is enrolled in more than 13 units?
- 20. What is the probability that if a student is working more than 20 hours per week, the student is enrolled in more than 13 units?
- 21. What is the probability that if a student is enrolled in more than 13 units, the student is working more than 20 hours per week?
- 22. What is the probability that if a student is enrolled in more than 13 units, the student is working fewer than 10 hours per week?
- 23. A BINGO card consists of a square containing 25 smaller squares, as shown in the figure. Five numbers from 1 to 15 are placed under the letter B, five numbers from 16 to 30 are placed under the letter I, four numbers from 31 to 45 along with the "free space" are placed under the letter N, five numbers from 46 to 60 are placed under the letter G, and five numbers from 61 to 75 are placed under the letter O.

When a regular Bingo game is played, the numbers from 1 to 75 are randomly called until someone's card has five numbers in a horizontal, vertical, or diagonal row. Suppose you are playing Bingo with the card in the figure.

- a) What is the probability that the first number called is on your card?
- b) What is the probability that G-59 is the first number called?
- c) What is the probability that the first number called is in your N row?
- d) Suppose the first number called is not on your card. What is the probability that the second number called is on your card?

В	Ι	N	G	0
14	21	39	47	68
7	22	44	53	64
3	16	FREE	60	61
1	29	42	46	75
13	18	31	59	71

Section 4.2 Expected Value and Odds

Expected value is the term used to describe the anticipated winnings from a contest or a game. It is used throughout the world to determine prizes in contests and premiums on insurance policies. In this section, we will show how an expected value can be calculated.

Suppose a game is played with one six-sided die. If the die is rolled and lands on 1, 2, or 3, the player wins nothing. If the die lands on 4 or 5, the player wins \$3. If the die lands on 6, the player wins \$12. The following table summarizes this information, including the probability of each situation.

Event	P(event)	Winnings
1, 2 or 3	3/6	\$0
4 or 5	2/6	\$3
6	1/6	\$12

If you play this game, how much can you expect to win? What is the expected value of this game?

Expected Value

To compute the expected value of a game, find the sum of the products of the probability of each event and the amount won or lost if that event occurs.

For this game, the expected value (E.V.) is given by $E.V. = \frac{3}{6} \times 0 + \frac{2}{6} \times 3 + \frac{1}{6} \times 12 = \3 .

An expected value of \$3 means that we would expect to win an average of \$3 for each game played. This means that if we played 1000 games, we would expect to win \$3000. However, as is true for probability, this is true only for a large number of games. If we played three games, although the expected winnings are \$9, we could win between \$0 and \$36.

Suppose the operator of the game charges \$1 to play the game. The amount won by the customer would then be reduced by \$1. By using the following table, we can compute the expected winnings for a player, including the cost of the game.

Event	P(event)	Winnings	
1, 2 or 3	3/6	-\$1	
4 or 5	2/6	\$2	
6	1/6	\$11	

For this game, the expected value is $E.V. = \frac{3}{6} \times (-1) + \frac{2}{6} \times 2 + \frac{1}{6} \times 11 = \2 .

As might be expected, a charge of \$1 to play the game reduces the expected value by \$1, from \$3 to \$2.

Example 1

A state lottery has 49 numbers, six of which are the winning numbers for a particular game. The cost of playing the lottery is \$1. To play the game, the player must pick six numbers. If a player picks three winning numbers, the payment is \$20. Similarly, picking four numbers pays \$100, picking five numbers pays \$10,000, and picking six winning numbers pays \$1,000,000. The probabilities and net winnings for each possiblity is given in the following table. Find the expected value of this game, including the cost of playing.

Event		
(number correct)	P(event)	Winnings
0	0.4359650	-\$1
1	0.4130195	-\$1
2	0.1323780	-\$1
3	0.0176504	\$19
4	0.0009686	\$99
5	0.0000184	\$9999
6	0.0000001	\$999,999

Solution: Calculating the expected value, we have

$$E.V. = (-1) \times 0.4359650 + (-1) \times 0.4130195 + (-1) \times 0.1323780 + 19 \times 0.0176504 + 99 \times 0.0009686 + 9999 \times 0.0000184 + 999,999 \times 0.0000001 \approx -\$0.27$$

By doing this calculation, we see that we expect to lose an average of 27ϕ every time we play the lottery. This means that the operators of the lottery expect to earn 27ϕ every time someone buys a ticket.

Example 2

A game is called **fair** if the expected value of the game is zero. Suppose a certain game is fair and costs \$3 to play. The probability of winning is 0.6, and the probability of losing is 0.4. How much should you win for the game to be fair?

Solution: Since the game is fair, we can set up an equation for the expected value of the game with E.V. = 0. We will use W to represent the prize for winning. Since it costs \$3 to play the game, the amount won if we win the game is W - 3 and the amount won if we lose the game is -3.

$$E.V. = (\text{amount won}) \times P(\text{winning}) + (\text{amount lost}) \times P(\text{losing})$$
$$0 = (W - 3) \times 0.6 + (-3) \times 0.4$$
$$0 = 0.6W - 1.8 - 1.2$$
$$0 = 0.6W - 3$$
$$0.6W = -3$$
$$W = 5$$

Therefore, since the game is fair, a player should receive \$5 if he wins. Note that this means a real winnings of only \$2, since it costs \$3 to play the game.

Insurance companies determine the premiums for a policy by examining the risk involved. The risk is calculated by looking at statistics involving the situations covered by the policy. A person in a high-risk category pays more for insurance than someone in a low-risk category does.

Example 3

The Lagomorph Insurance Company has a customer, Mr. Roger Abbit, who holds a \$250,000 fire insurance policy on his art collection. The company estimates that there is a 1% chance that the art will be destroyed by fire. If the insurance company tries to maintain an expected value of \$200 on each policy, what should Mr. Abbit's premiums be?

Solution: If a = amount of the premiums, then a - 250,000 is the amount the insurance company will lose if Roger's art collection is destroyed by fire. Since the company estimates that there is a 1% chance of loss by fire, there is a 99% chance that there will be no loss.

 $E.V. = (\text{amount lost}) \times P(\text{losing}) + (\text{amount paid in premiums}) \times P(\text{not losing})$ $200 = (a - 250,000) \times 0.01 + a \times 0.99$ 200 = 0.01a - 2500 + 0.99a 200 = a - 25002700 = a

This means that the insurance Company should charge \$2700 for this policy.

Determining Odds of Single Events

In many situations, rather than probability, the odds of an event are given. The **odds** for any event can be thought of as a ratio of the number of desired outcomes to the number of outcomes that are not desired, such as wins to losses or successes to failures.

Example 4

The odds of Giacomo winning the Kentucky Derby in 2005 were 1 to 50. Explain what this means and give the probability that Giacomo would win.

Solution: Odds of 1 to 50 indicate that for every 1 time Giacomo wins, he loses 50 times. Since probability is the ratio of the number of successes to the total number, finding the probability requires knowing the total. On the basis of odds, we can think of Giacomo

running 51 (1 + 50) times. Thus, the probability of Giacomo winning was $\frac{1}{51}$.

As seen in Example 4, like probability, odds are also used to measure the likelihood of an event occurring. We were able to convert from odds to probability by thinking about the meaning of both odds and probability. While we feel that this is a straight-forward way to convert from one to the other, there are also formulas to do the conversions.

To convert from odds to probability:

If the odds for an event E are a to b (a:b), then

$$P(E) = \frac{a}{a+b}.$$

To convert from probability to odds:

If the probability of an event *E* is *P*(*E*), then the odds, *O*(*E*), are given by $O(E) = \frac{\text{number of ways an event can occur ($ *desired* $)}}{\text{number of ways an event fails to occur ($ *not desired* $)}} = \frac{P(E)}{1 - P(E)}$

Example 5

Suppose the San Francisco Giants have a 10% chance of winning the National League pennant. What are the odds of the Giants winning the pennant? What are the odds of the Giants not winning the pennant?

Solution: Since the probability is 10%, P(E) = 0.10. The odds of the Giants winning the pennant are

$$O(E) = \frac{0.10}{1 - 0.10} = \frac{0.10}{0.90} = \frac{1}{9} = 1:9$$

The odds of the Giants not winning the pennant are 9:1.

Example 6

The chairman of a corporation states in a press conference that the odds are 10 to 1 that the company will reach \$1,000,000,000 (one billion) in sales in the coming year. What is the probability that the corporation will reach one billion dollars in sales this year?

Solution: Since the odds are 10 to 1, use a = 10 and b = 1 in the formula that converts odds to probability. Thus,

$$P = \frac{a}{a+b} = \frac{10}{10+1} = \frac{10}{11} \approx 91\%$$

There is a 91% chance of reaching one billion dollars in sales.

House Odds

The odds for various events as given by casinos, racetracks, and lotteries are called **house odds**. They give the odds against an event occurring. For example, before the 2006 baseball season, a Las Vegas casino stated that the odds for the New York Yankees to win the World Series were 7:2. This really meant that the odds against the Yankees winning the World Series were 7 to 2 or the probability of the Yankees winning was 2/9. House odds are given in such a manner as to facilitate the betting of money that might accompany the event. Odds of 7 to 2 would mean that for every \$2 you bet on the Yankees, you would win \$7 if the Yankees won the World Series. The odds for the Tampa Bay Devil Rays to win the World Series in the same year were 100:1. This would mean that if Tampa Bay won the World Series, you would win \$100 for every \$1 you bet. Odds are established to balance the amount won with the likelihood

that the event occurs. Less probable events have higher house odds and greater potential winnings if the event occurs.

Example 7

- The oddsmakers have stated that Jhun's odds are 7 to 1 in the title fight with Chavez.
- a) What is the probability that Jhun wins the match?
- b) If you bet \$50 on Jhun, and Jhun beats Chavez, how much money will you win?

Solution:

a) The house odds of 7 to 1 indicate that Jhun's odds of losing are 7 to 1. This means that the odds of Jhun winning are 1 to 7. Using a = 1 and b = 7 in the formula that converts odds to probability, we find that the probability that Jhun wins the fight is

$$P = \frac{a}{a+b} = \frac{1}{1+7} = \frac{1}{8}.$$

b) The house odds for Jhun of 7 to 1 indicate that if Jhun wins, each \$1 bet on Jhun will generate \$7 in winnings. Thus, if you bet \$50, you will win $7 \times $50 = 350 .

Example 8

An American roulette wheel has thirty-eight compartments around its circumference. Thirty-six compartments are numbered from 1 to 36, half of them being colored red and half colored black. The remaining two compartments, numbered 0 and 00, are colored green. A ball is spun and lands in one of the compartments. The odds for betting on a single number on an American roulette wheel are given by casinos as 35:1.

- a) What do the odds mean in terms of the amount won when \$1 is bet on a single number?
- b) On the basis of the actual roulette table and not the odds, what are the odds of winning a single bet?
- c) Are the odds posted by the casino fair to the players?

Solution:

- a) House odds of 35:1 indicate that for every winning dollar bet on a single number, the bettor wins \$35.
- b) If *E* is the event that the single number comes up,

$$P(E) = \frac{1}{38}, \quad O(E) = \frac{P(E)}{1 - P(E)} = \frac{\frac{1}{38}}{1 - \frac{1}{38}} = \frac{\frac{1}{38}}{\frac{37}{38}} = \frac{1}{37}.$$

c) The odds against the single number coming up are 37:1. The house odds are not correct. Since the casino is paying out \$2 less than is mathematically correct, the odds are not fair to a player. The discrepancy in the odds gives the casino a margin of profit.

§4.2 Explain \rightarrow Apply \rightarrow Explore

Explain

- 1. What is expected value and how is it calculated?
- 2. In terms of expected value, when is a game considered to be fair?
- 3. What does it mean for the expected value of a \$1 slot machine to be -\$0.15?
- 4. What does it mean for the expected value of a \$1 lotto game to be \$0.15?
- 5. Why is your expected value for games of chance in a gambling casino negative? Explain.
- 6. What are odds?
- 7. What are house odds?
- 8. What do house odds tell you about money being bet on an event?
- 9. How do you convert from probability to odds?
- 10. How do you convert from odds to probability?

Apply

- 11. In a certain game, the probability of winning is 0.3, and the probability of losing is 0.7. If a player wins, the player will collect \$50. If the player loses, the player will lose \$5. What is the expected value of this game? If the game is played 100 times, what are the expected winnings (or losses) of the player?
- 12. In a game of dice, the probability of rolling a 12 is 1/36. The probability of rolling a 9, 10, or 11 is 9/36. The probability of rolling any other number is 26/36. If the player rolls a 12, the player wins \$5. If the player rolls a 9, 10, or 11, the player wins \$1. Otherwise, the player loses \$1. What is the expected value of this game? If the game is played 100 times what are the expected winnings (or losses) of the player?
- 13. In a game of dice, the probability of rolling a 12 is 1/36. The probability of rolling a 9, 10, or 11 is 9/36. The probability of rolling any other number is 26/36. If the player rolls a 12, the player wins \$8. If the player rolls a 9, 10, or 11, the player wins \$2. If the game is fair, how much should the player lose when the player rolls any number less than 9?
- 14. Many charities and other organizations use lotteries or other similar marketing devices to acquire funds. Many states' laws require the odds of winning the various prizes to be posted on the back of the tickets or in some other conspicuous spot.
 - a) Fancy That Poultry magazine is running a contest with the following odds and prizes.

Odds	Prize
1 to 49	\$10 (in back issues)
1 to 9999	\$50 (in poultry feed)
1 to 99,999	\$2000 (in rare ducks)

If tickets are free, find the expected value of a "winning ticket."

- b) Is the drawing worth the price of a first-class stamp?
- c) At what postage rate would the drawing be considered fair?
- 15. The odds of an event are 3 to 2. What is the probability of the event?
- 16. The odds of an event are 1 to 3. What is the probability of the event?
- 17. The probability of an event is 25%. What are the odds of the event?

- 18. The probability of an event is 1/10. What are the odds of the event?
- 19. The house odds of a game are 10 to 1. What is the probability that the house wins? What is the probability that you win?
- 20. The house odds of a game are 20 to 1. What is the probability the house wins? What is the probability you win?

Explore

21. At many bingo parlors, the operators sell pull tabs, which are very similar to slot machines. Each pull-tab card has rows of symbols that are covered by paper tabs. If the paper tabs are removed and three of the same symbols are in a straight line, you win a designated amount. A summary of a typical pull-tab game in which each card costs \$0.50 is shown below.

Symbols	Probability	Amount Won
Three diamonds	2/2783	\$149.50
Three rubies	2/2783	74.50
Three pearls	4/2783	9.50
Three coins	20/2783	2.50
Three stars	250/2783	0.50
Three moons	400/2783	0
Any other combinations	2105/2783	-0.50

Determine the expected value of this pull-tab game.

- 22. In the game of roulette, players bet that a ball will land on a certain number. A player can choose any number from 1 through 36, 0, or 00. It costs \$1 to play the game. If the player correctly selects the number, the \$1 is returned, and the player receives an additional \$35. What is the expected value of this game? Suppose a casino has 100 players, each of whom plays ten times each hour for 24 hours. Each player bets \$1. What is the casino's expected profit?
- 23. Dennis is in charge of designing a game for the school fund raiser. Participants will be paying \$2 for each game. There will be three prizes. The lowest has a value of \$0.50, the second has a value of \$1, and the first prize has an undetermined value. The probability of winning the lowest prize is 0.35, the probability of winning the second prize is 0.15, and the probability of winning the first prize is 0.01. The probability of not winning any prize is 0.49. If the school wants an expected value of \$1 per ticket, what should Dennis choose as the value of the first prize?

24. In the game of keno, 80 numbers are displayed on a board. Twenty of these numbers are chosen to be the winning numbers. Suppose a person has selected five numbers on a keno card. The probabilities and winnings are as follows.

Event	Probability	Amount Won
0 winning numbers	0.227184	-\$1
1 winning number	0.405686	-\$1
2 winning numbers	0.270457	-\$1
3 winning numbers	0.083935	\$1
4 winning numbers	0.012092	\$10
5 winning numbers	0.000645	\$250

- a) What is the expected value of the game?
- b) If the game costs \$1, how much money should the player expect to win or lose after playing 1000 games?
- 25. A television game show contestant has current prizes worth \$12,500. If the contestant participates in the next round of competition, he will have \$50,000 if he wins and \$0 if he loses. Assuming the game show is fair (E.V. = 0), what is the contestant's probability of winning \$50,000?
- 26. If the probability of you making a three-point shot in basketball is 0.15,
 - a) What is the probability that you will not make a three-point shot?
 - b) What are the odds of you making a three-point shot?
 - c) What are the odds of you not making a three-point shot?
- 27. In the television show, "Deal, No Deal," a contestant has one of four briefcases with one of the following prizes: \$400, \$1000, \$500,000, or \$1,000,000. He is offered \$357,000 to give up his case. Use expected values to determine if he should accept the \$357,000 deal or keep his case ("No Deal").
- 28. If there is 75% chance that you will get up when the alarm rings,
 - a) What is the probability that you will not get up when the alarm rings?
 - b) What are the odds of you not getting up when the alarm rings?
 - c) What are the odds of you getting up when the alarm rings?

Section 4.3 Statistical Graphs

Statistics are numerical data assembled in such a way as to present significant information about a subject. You come into contact with statistics on a daily basis. Credit bureaus use your payment history to determine your credit rating. Schools use standardized tests to determine student placement. Executives accumulate and organize data for a presentation. Government officials present findings regarding pollution, demographics, and economics. Magazines and newspapers contain articles with charts and data in almost every issue. In this section, we discuss methods of graphically presenting data.

The Data

When a study or a poll is done, much of the information is in numerical form. Merely listed, the results of a survey can be overwhelming. Suppose 100 people respond to a survey with ten questions, each with numerical answers. There will be 10 numbers from each survey, giving a total of 1000 numbers. Most people will not be able to make any pertinent observations when the data are in this form. The following will show how to overcome such a difficulty.

Displayed here are the test scores of 70 students who have taken a math placement test. The test has a possible low score of 0 and a possible high score of 25.

12	23	25	5	9	5	24	20	3	16
14	14	15	21	2	18	13	11	18	14
22	16	17	15	19	23	14	23	2	21
11	8	7	16	11	10	19	14	21	10
24	16	11	22	20	14	17	22	7	10
11	13	18	9	6	15	4	12	19	25
6	23	20	13	9	7	15	16	11	15

As we look at the data, there does not appear to be any pattern, nor can we come to any conclusions about the test scores. To examine the data more carefully, one technique is to put the data into categories.

When data are divided into categories, the different categories are called **classes**. The number and size of the classes we choose are determined by the goals of the investigation. Often, studies have predetermined classes. For example, many college courses use the standard groupings 90 - 100 for an A, 80 - 89 for a B, 70 - 79 for a C, 60 - 69 for a D, and 0 - 59 for an F. As another example, economic studies concerned with household income have classes such as "below the poverty level (less than \$18,400 income per year)," and "lower middle income (between \$18,400 and \$31,000)."

A second criterion for choosing the classes is common sense. It would not be very wise to pick only one class because all the data will fall into that class. On the other hand, creating 30 classes for our data would mean that some of the classes would be empty and most of the classes would contain only a few scores. Finally, if we want the number of items in a class to be meaningful, the size of each class should be the same.

For our test data, we do not have any predetermined classes. Since the data vary from 0 through 25, it is convenient to pick five classes. This number was chosen because it allows us

0.2

to create nearly equal size classes. The classes are 0 through 5, 6 through 10, 11 through 15, 16 through 20, and 21 through 25.

Now that we have chosen the classes, the next step is to determine the **frequency**, or the number of data values in each class. To do this, we have created the following chart, called a frequency distribution. The column labeled "Tally" is used to count the number of scores that fall into a class. The column labeled "Frequency" contains the number of tallies for that class. The entries in the column labeled "Relative Frequency" are found by dividing the frequency for each class by the total number of items, in this case 70. The final column is called "Percentage Frequency". Its entries are found by multiplying the relative frequency by 100. They are the percentages of the data that fall into each category. With the information presented in this way, we can see that of the 70 test scores, 22 were in the class 11-15. This group comprised 31% of those who took the test.

Frequency Distribution						
			Relative	Percentage		
Class	Tally	Frequency	Frequency	Frequency		
0-5	₩ Ι	6	0.09	9%		
6-10	HH HH II	12	0.17	17%		
11-15	++++ +++ +++	22	0.31	31%		
16-20	++++ +++ 1	16	0.23	23%		
21-25	++++ +++	14	0.20	20%		

Having the data organized in this fashion is a great improvement, but it is still a bit intimidating. Many people do not like to look at tables of numbers. Information can often be presented with greater impact if it is in graphical form. For our discussion, we will examine three of the many types of graphical representations. The first of these is called a **bar graph**. A bar graph uses one axis to represent the different classes, while the other axis is used to indicate the frequency for each class.

Bar graphs may be oriented horizontally or vertically and may use either the frequencies or the relative frequencies as the lengths of the bars. The choice of display is left to the person creating the graphs. With a bar graph, it is easy to see which class contains the largest number of test scores. The numbers at the ends of the bar permit the reader to determine the actual number of scores in each class. Two different bar graphs of the data are shown.



A second way in which information can be displayed is with a **line graph**, also called a frequency polygon or broken line graph. In a line graph, the bars of a bar graph are replaced with dots that are connected by line segments. A line graph that represents the frequency of scores on the math placement test is shown in the following figure.

A line graph allows you to quickly see the increases and decreases in the number of scores in each class. For example, you can see that



there is a large jump between the number that scored from 6 to 10 and the number that scored from 11 to 15. On the other hand, you can see that there is only a slight decrease from the number that scored from 16 to 20 and the number that scored from 21 to 25.

A third way in which information can be presented is through the use of a pie chart, also called a circle graph. A **pie chart** is a circular diagram divided into sectors (wedgeshaped pieces), where the areas of the sectors are used to represent percentage frequencies. Each sector represents a part or percentage of a whole. The percentage frequency determines the angle of the sector, the entire circle representing 100%. The percentage frequencies of the math placement test can be dis-



played in the pie chart that follows. From the pie chart, you can visually determine the relative size of each class. You can quickly see which class had the largest number of scores and which class had the smallest number of scores.

Which Type of Graph Should Be Used?

Since there are three basic types of statistical graphs, which one should be used for a given set of data? Depending on what you are trying to emphasize, the data can be displayed with any of these three statistical graphs. If you want to clearly show which category contains the largest or the least number of items, a bar graph is effective. If you want to emphasize the patterns of change (rise and fall) of various categories, a line graph is useful. If you want to show how a whole is divided into parts, then a pie chart is the best one to use.

Making Graphs

Today's computers accomplish the drawing of these graphs by using chart-drawing software.

Often these abilities are incorporated into large software packages, such as spreadsheet programs and word processors. However, it might be necessary to create a graph without the aid of a computer. The remaining part of this section describes how to draw these basic statistical graphs.

Suppose we want to look at a breakdown of how people are employed in the United States according to occupation and sex. According to statistics published by the U.S. Department of Labor, the employment classification for each gender is given in the following table.

Occupation	% Males	% Females
Executive, administrative, management	15.0%	14.2%
Professional specialty	13.6%	18.0%
Technical, sales, administrative support	19.7%	40.0%
Service occupations	9.9%	17.4%
Precision production, crafts	18.6%	2.1%
Operators, fabricators, laborers	19.3%	7.1%
Farming, forestry, fishing	3.8%	1.1%

The data can be represented by either a bar graph or pie chart.

Bar Graphs

To make a bar graph, determine the following:

- 1. The categories to be placed along the vertical and horizontal axes
- 2. The scale used to encompass the numerical data
- 3. The style of bars used and the length of each bar

One possible way to draw the bar graph comparing occupation data is to have the vertical axis measure the percent and the horizontal axis contain the occupation. Since the data consists of percentages that range from 1% to 40%, by letting the vertical scale be $\frac{1}{2}$ in. for every 10%, we could fit the graph in a reasonable area. Since each 10% will be represented by 0.5 in., to find the length of each bar, we could set up a proportion such as

$$\frac{0.5 \text{ in.}}{10\%} = \frac{L \text{ in.}}{P\%}.$$

For 15% we get:

$$\frac{0.5 \text{ in.}}{10\%} = \frac{L \text{ in.}}{15\%}$$
$$L = \frac{0.5}{10\%} \times 15\% = 0.75 \text{ in.}$$

Finally, since we are comparing the occupations of males versus females, we can use light gray bars for males and dark gray bars for females. Using vertical bars that fit the data, we get the following statistical graph.



The bar graph allows us to visually analyze the data. By simply looking at the bar graph, we can make conclusions about percentage of a gender in a particular field. For example, 40% of women are employed as technical, sales, and administrative support staff, compared to only 20% of men.

Pie Charts

To make a pie chart, determine the size of each sector that will be used to represent the percentage frequency. Since there are 360 degrees around the center of a circle, the angle for each sector is found by multiplying the percentage by 360° . For example, to find the angle used to represent the 15% of men employed in management, multiply $0.15 \times 360^\circ = 54^\circ$.

Using a protractor to measure the desired angles, each circle is divided into seven sectors. Using a different color for each occupation and including labels, we get the following pie charts for the occupations of each gender.

As we did with the bar graphs, we can draw conclusions about the occupations of each sex. For example, we can see that a large percentage of women are employed in technical, sales, and administrative support positions.



A quick look through magazines and newspapers will convince you that graphs are a common way of displaying information. While software packages often have many different types of graphs, the three types of graphs presented here give you a good basis for understanding all of them.

§4.3 Explain \rightarrow Apply \rightarrow Explore

Explain

- 1. What are statistics?
- 2. What is a bar graph?
- 3. What is the difference between the frequency, relative frequency, and percentage frequency distribution?
- 4. What is a line graph?
- 5. What is a pie chart?
- 6. What are the three major steps in making a bar graph?
- 7. How is the angle of each sector of a pie chart determined?
- 8. When would it be effective to display data using a bar graph? line graph? pie chart?

Apply

- 9. For the data 2, 4, 6, 1, 11, 9, 5, 3, 7, 6:
 - a) Arrange the data into a frequency distribution with three classes.
 - b) Draw a bar graph, line graph, and pie chart.

- 10. For the data 2, 4, 3, 7, 2, 8, 6, 3, 7, 5:
 - a) Arrange the data into a frequency distribution with four classes.
 - b) Draw a bar graph, line graph, and pie chart.
- 11. For the data 10, 12, 14, 13, 17, 12, 18, 16, 13, 17, 15, 17:
 - a) Arrange the data into a frequency distribution with five classes.
 - b) Draw a bar graph, line graph, and pie chart.
- 12. For the data 20, 29, 21, 21, 28, 28, 23, 23, 23, 26, 26, 26, 25, 25, 25, 25, 25;
 - a) Arrange the data into a frequency distribution with five classes.
 - b) Draw a bar graph, line graph, and pie chart.

Explore

- 13. The bar graph (from statistics compiled by the U.S. Department of Labor) shows the U.S. Consumer Price Index (CPI). The CPI represents the relative costs of goods. Using 1967 as the base year, something that cost \$100 in 1967 cost \$30 in 1915 and \$512 in 2000.
 - a) If a washing machine cost \$350 in 1967, what would the machine cost in 2000?
 - b) If a teacher earns \$40,000 per year in 2000, what would her pay have been in 1915?
 - c) Look up the CPI for 2001, 2003, and 2005 and add these values to the chart. What conclusions can you make?



Consumer Price Index 1915 - 2000

- 14. The following bar graph shows the number of DVD players sold in the years 1997-2005.
 - a) How many more DVD players were sold in 1999 than in 1998?
 - b) Find the percentage of the total DVD players sold in each of the years 1997-2005 and create a pie chart from the information. Use the sum of the DVD players sold in the nine years for the total DVD players sold.
 - c) Look at the DVD player sales for 2001 through 2005. What conclusions can you make about sales of DVD players? Give a reason why this may be happening.



DVD Player Sales 1997 - 2005

- 15. Look up the average price of unleaded gas for years between 1970 and the present. You should use numbers that are adjusted for inflation. Look at the graph and see what conclusions you can draw.
- 16. Look up the average annual health care costs for a typical family for years between 1970 and the present. You should use numbers that are adjusted for inflation. Look at the graph and see what conclusions you can draw.

Section 4.4 Analyzing Data

There are two primary computations that are done when working with statistical data, finding an "average" and finding the standard deviation. Finding an average is a computation that requires only arithmetic skills. It has, however, very important uses in real-world problems. An average allows us to find one value that represents the middle of a set of data. As we shall see, however, the middle of a set of data can be described in different ways. Collectively, these ways of describing the middle value are called measures of **central tendency**. While an average describes the center of a set of data, the standard deviation describes how much the data spreads out. In this section, we introduce three measures of central tendency and two measures for the spread of the data.

The Mean

The **mean** or **arithmetic** (pronounced ar-ith-met'-ic) **mean** of a set of values is what most people are referring to when they say "average." It is found by adding up all the data and then dividing by the number of data values. Statisticians use the symbol μ (mu) for the mean and define it with the following formula. In the formula, $\sum x$ is simply a mathematical abbreviation for finding the sum of the data and *n* is the number of data values.

mean =
$$\mu = \frac{\text{sum of the data values}}{\text{number of data values}} = \frac{\Sigma x}{n}$$

Example 1

Find the mean of the numbers 3, 6, 8, 5, 4.

Solution:
$$\mu = \frac{3+6+8+5+4}{5} = \frac{26}{5} = 5.2$$

Example 2

Find the mean of 1, 1, 1, 1, 3, 3, 5, 5, 6, 6, 6, 6.

Solution: Rather than merely adding all these numbers, it is more convenient to multiply each value by its frequency and then add. This total is then divided by the total frequency. In other words, rather than doing the computation

$$\mu = \frac{1+1+1+1+3+3+5+5+6+6+6+6+6}{13} = \frac{50}{13} \approx 3.85$$

it is easier to do the calculation

$$\mu = \frac{4(1) + 2(3) + 2(5) + 5(6)}{13} = \frac{50}{13} \approx 3.85$$

A common use of the above calculation is to compute a **weighted average**. A weighted average is the mean of a group of numbers in which certain values have more importance, or weight, than do other values. When computing a weighted average, we use the weights (*w*) of each value as the multipliers and the sum of the weights as the divisor.

Weighted Average =
$$\mu = \frac{\text{sum(weights \times values)}}{\text{sum of the weights}} = \frac{\sum(w \cdot x)}{\sum w}$$

An example of this type of situation can be seen in the next problem.

Example 3

A student is trying to calculate his grade in an English class. His scores are as follows.

Midterms82Homework87Final92

The course syllabus says that the midterms count for 60% of the grade, homework for 10% of the grade, and the final for 30% of the grade. Compute the student's score in the class.

Solution: Using the percentages as the weights of the scores in each category, we have

$$\mu = \frac{0.60(82) + 0.10(87) + 0.30(92)}{0.60 + 0.10 + 0.30} = \frac{85.5}{1.00} = 85.5$$

The remainder of this section will assume that you have a calculator that has a function to compute the mean of a set of data. For help with your calculator, read your calculator manual or ask your instructor for help.

The Median

Example 4

Five houses are listed for sale in a real estate broker's advertisement. The prices are \$269,000, \$256,000, \$249,000, \$235,000, and \$749,000. Find the average price of the houses listed by the agent.

Solution: Calculating the mean, we get $\mu = $351,600$. Notice that the mean is higher than all but one of the house prices. The reason is that the price of the most expensive house is distorting the results. The mean price of these five houses is \$351,600, but the mean does not accurately reflect the typical or "average" price.

To solve this difficulty, we will use a measure of central tendency called the median. The **median** is found by listing the data in increasing order and choosing the middle value. If there is an even number of items, the median is found by taking the mean of the two middle values. The median is frequently used when there are a few extreme values in the data that will greatly affect the value of the mean.

Example 5

Find the median of the prices listed in Example 4.

Solution: Listing the data in increasing order gives

\$235,000, \$249,000, \$256,000, \$269,000, \$749,000.

The median price is \$256,000 because this is the middle value. This value is much closer to what we would call the "average" price of these five houses. Also notice that

increasing the price of the most expensive house will have no effect on the value of the median.

Example 6

Find the median of the values 0, 3, 4, 9, 16, 90.

Solution: The values are listed in order, so we need only find the middle value. Since there is an even number of values, the median is found by calculating the mean of the center pair of values.

$$median = \frac{4+9}{2} = 6.5$$

The Mode

The **mode** of a set of data is the value that occurs most frequently. It is the only measure of central tendency that can be used with nonnumerical as well as numerical data. For example, if a design consultant wanted to determine the most popular color for exterior house paint, he or she could conduct a survey to find out how many people liked each color. It would not be possible to use the mean or median but the color with the most votes, the mode, could determine the most popular color.

Example 7

Determine the mode of the following sets of values.

- a) 1, 2, 2, 3, 4, 6, 6, 6, 8, 9
- b) 1, 2, 2, 3, 4, 6, 6, 7, 8, 9
- c) 1, 2, 2, 3, 4, 4, 6, 6, 8, 9

Solution:

- a) In the first group of data, the number 6 appears three times, and all the other numbers appear at most twice. Therefore, the mode is 6.
- b) In the second group, the numbers 2 and 6 both appear twice. No other number appears more than once. In a situation like this, there are two modes: 2 and 6. When a set of data has two modes, it is called **bimodal**.
- c) In the third set of data, the numbers 2, 4, and 6 all appear twice. When more than two values have the highest frequency, we say that the set of data does not have a mode.

Example 8

In 1990, there were 66,090,000 families in the United States. Of these families, 33,801,000 had no children, 13,530,000 families had one child, 12,263,000 had two children, 4,650,000 families had three children, and 1,846,000 families had four or more children.

- a) Determine the modal number of children per family.
- b) Determine the mean number of children per family.

Solution:

- a) Since the class with the highest frequency is the class with no children, the modal number of children is 0.
- b) Finding an accurate mean of the data will not be possible. Since the final category consists of families with four or more children, we do not know the actual number

of children in these families. However, if we use 4 as the number of children for the last group, we can use the method of finding the mean for grouped data to give an estimate.

 $\mu = \frac{33,801,000(0) + 13,530,000(1) + 12,263,000(2) + 4,650,000(3) + 1,846,000(4)}{33,801,000 + 13,530,000 + 12,263,000 + 4,650,000 + 1,846,000}$ $= \frac{59,390,000}{66,090,000} \approx 0.90$

Since the number of children in the last class may be greater than four per family, we can say that the mean number of children per family is at least 0.90.

Measures of Dispersion

Suppose the mean, mode, and median for two sets of data are identical. Does this suggest that the data are the same? Consider the two sets of data: 1, 1, 100, 100, 100, 199, 199 and 99, 99, 100, 100, 100, 101, 101. For each set, the mean, mode, and median all equal 100, yet the data are not the same. Not only are the data not the same, the first set has values that range between 1 and 199, whereas the second set is closely clustered around 100. From this, we can see that we need more tools to help describe a distribution of numbers. Collectively, the tools used to do this are called measures of **dispersion**. A measure of dispersion will provide a tool to determine the extent to which the data in a set differ from a central value. In this section, we will discuss two of these measures, range and standard deviation.

The Range

The range of a set of values is the difference between the highest and lowest values in the set.

Example 9

Find the range of each of the following sets of numbers.

- a) 1, 1, 100, 100, 100, 199, 199
- b) 99, 99, 100, 100, 100, 101, 101

Solution:

- a) The range of the first set of data is 199 1 = 198.
- b) The range of the second set of data is 101 99 = 2.

We can see that the range will provide some help in analyzing the difficulty mentioned in the introductory remarks. However, it will not completely solve the problem. Consider the following two sets of data: 1, 1, 1, 100, 199, 199, 199 and 1, 100, 100, 100, 100, 100, 100, 199. In both sets, the range is 198, but the first set of data has most of its values at the extreme ends, whereas most of the data in the second set have the value of 100. Only two of the numbers are at the extremes. What is needed is a method to determine the average of the distance between each data value and the mean. If this average is high, then the data are spread out. If the average is low, then the data are clustered together.

Standard Deviation

The measure that we will use to determine how closely the data are clustered around the mean is called the **standard deviation**. The standard deviation is the square root of the average of the squares of the differences between the data values and the mean. The standard deviation is

$$\sigma = \sqrt{\frac{\text{sum}(\text{data values} - \text{mean})^2}{\text{number of data values}}} = \sqrt{\frac{\sum (x - \mu)^2}{n}}$$

Example 10

Find the standard deviation of the numbers 4, 6, and 11.

Solution: To find the standard deviation, first find the mean.

$$\mu = \frac{4+6+11}{3} = 7$$

Next, subtract the mean from each of the data values and then square the results. Using the formula gives the following.

$$\sigma = \sqrt{\frac{(4-7)^2 + (6-7)^2 + (11-7)^2}{3}}$$
$$= \sqrt{\frac{9+1+16}{3}}$$
$$= \sqrt{\frac{26}{3}} \approx 2.94$$

Although this process might not seem too difficult, performing the standard deviation calculation by hand is very tedious. Although algebraic techniques exist to simplify this process, the ready availability of calculators eliminates the need for tedious calculations. As a result, for the remainder of this chapter we will assume that you have a calculator that has a standard deviation function. For help with your calculator, read your calculator manual or ask your instructor. If you haven't done so already, use the standard deviation function on your calculator to find the standard deviation of the numbers given in Example 10.

How Is Standard Deviation Used?

The standard deviation of data is a measure of how much the data spreads out from the mean but, so far, we have no real feeling for what standard deviation indicates or how it can be used. In the remainder of this section, we discuss the meaning of standard deviation in terms of a frequency distribution.

Example 11

Consider the following sets of test scores from two algebra classes.

Morning Class	Evening Class	
78, 65, 83, 91, 98, 25, 67, 88, 81, 77,	34, 87, 81, 93, 99, 24, 77, 62, 98, 100,	
53, 76, 80, 72, 75, 69, 64, 62, 85, 93,	57, 31, 81, 72, 61, 59, 68, 74, 77, 94,	
70, 44, 85, 73, 75, 63	56, 71, 70, 81, 78, 83, 25, 94, 31	



- a) Find the mean and standard deviation of each class.
- b) Use the bar chart to draw conclusions about the meaning of standard deviation.

Solution:

a) Using a calculator to find the mean and standard deviation of each class, we have the following.

Morning Class: $\mu \approx 73$, $\sigma \approx 15.32$ Evening Class: $\mu \approx 70$, $\sigma \approx 22.05$

b) Notice that the scores for the evening class (light colored bars) are more spread out that the scores for the day class, especially at the left end of the bar chart. This corresponds with the values given by the standard deviation. The standard deviation of the evening scores is larger than the standard deviation of the morning scores.

The Normal Distribution

Standard deviation can be better understood if we consider a frequency distribution whose line graph has a "bell" shape where frequencies start low, increase to a maximum, and decrease to a low frequency. Such distributions are common in many sets of data, for example, in standardized tests, in analysis of heights and weights, and in games of chance such as dice. The normal distribution is what is referred to when an instructor says you are being graded on a "curve". The mean is in the middle of the curve and one-half of the scores are below the mean and one-half the scores are above the mean. The standard deviation tells you how tightly the data is clustered around the mean. When the data is bunched together near the mean, the bell-shaped curve is steep and the standard deviation is small. When the data is spread apart, the bell shaped curve is relatively flat and the standard deviation is large. In any case, the normal distribution with its bell-shaped curve has the properties shown below.



- 1. About 68% of all values fall within 1 standard deviation of the mean.
- 2. About 95% of all values fall within 2 standard deviations of the mean.
- 3. About 99.7% of all values fall within 3 standard deviations of the mean.

Let's see what this means by examining another example.

Example 12

Suppose an IQ (Intelligence Quotient) test has a normal distribution with a mean of 100 and a standard deviation of 15. Determine the range of scores 1, 2, and 3 standard deviations from the mean and explain what that indicates.

Solution:

Since the mean, $\mu = 100$, and the standard deviation, $\sigma = 15$, then

- 1. 68% of IQ scores fall between 100 + 15 = 115 and 100 15 = 85.
- 2. 95% of IQ scores fall between 100 + 2(15) = 130 and 100 2(15) = 70.
- 3. 99.7% of IQ scores fall between 100 + 3(15) = 145 and 100 3(15) = 55.

The results of Example 12 can lead us to some interesting conclusions.

- 2.5% have an IQ score greater than 130.
- 0.15% have an IQ score greater than 145.
- If 1000 people took the IQ test, we would expect 68% or 680 of them to be clustered around the mean of 100, that is, have a score between 85 and 115.

§4.4 Explain \rightarrow Apply \rightarrow Explore

Explain

- 1. What is a measure of central tendency?
- 2. What is the mean for a set of data and how is the mean determined?
- 3. What is the weighted average? When is it used?
- 4. What is the median for a set of data?
- 5. What conditions in the data would make the median an effective measure of central tendency?
- 6. What is the mode for a set of data?
- 7. For what kind of data would the mode be an effective measure of central tendency?
- 8. What is a measure of dispersion?
- 9. What is the range of a set of data?
- 10. What is the standard deviation of a set of data?
- 11. You are comparing data from two similar experiments that have a mean of 34.5. Experiment A has a standard deviation of 10.1, and experiment B has a standard deviation of 5.5. What does this tell you about the data in the experiments?
- 12. If a set of test scores has a standard deviation of zero, what can be said about the scores?
- 13. Describe the characteristics of a normal distribution.
- 14. How are percentages and standard deviations used in interpreting the values of a normal distribution?

Apply

- 15. For the set of values 2, 4, 7, 2, 1, 8, 9, 10, 9, 6:
 - a) Find the mean.
 - b) Find the median.
 - c) Find the mode.
 - d) Find the range.
 - e) Find the standard deviation.
 - f) In the data, suppose the 1 is replaced by 4, and the 10 is replaced by 6. Without doing any calculations, what effect does this have on the standard deviation?
- 16. A student had seven 100-point tests during the semester. The scores on his tests were 90, 92, 59, 65, 94, 73, and 94. The instructor has given him the option of determining his semester grade by using the mean, median, or mode of his scores. Which would be the fairest selection? Explain.
- 17. A student is computing her cumulative grade point average (GPA). During six semesters, she has received 5 *C*'s, 11 *B*'s, and 8 *A*'s. If a grade of *C* is worth 2 points, a *B* worth 3 points, and an *A* worth 4 points, and all classes have the same number of units, compute the student's GPA. (Hint: Use the number of grades as the weights.)
- 18. In a large lecture class of 200 students, the instructor grades on the curve (normal distribution). He states that the average on the test was 72% and the standard deviation was 12. What range of scores fell within 1 standard deviation of the mean and how many student's scores were in that region?

- 19. In the previous problem, what range of scores fell within 2 standard deviations of the mean and how many student's scores were in that region?
- 20. According to the National Center for Health Statistics children's weights are normally distributed and newborn boys have an average weight of 7.8 pounds with a standard deviation of 1.1 lbs. Using this, determine the range of weight for 95% of newborn boys clustered around the mean.
- 21. According to the National Center for Health Statistics children's heights are normally distributed and 12 month old girls have an average length of 29.1 inches with a standard deviation of 1.1 inches. Using this, determine the range of length for the 68% of 12 month old girls clustered around the mean.

Explore

- 22. An accounting firm plans to buy a large number of cartridges for laser printers. The cartridge is available from two different suppliers. The first supplier says that the expected lifetime is 3000 pages with a standard deviation of 100 pages. The second supplier says that its cartridge has an expected lifetime of 3000 pages with a standard deviation of 400 pages. If you are the buyer for the accounting firm, which supplier would you choose? Explain your reasoning.
- 23. Find the salaries of the top ten highest-paid athletes over the most recent season for one of the following major league sports: the National Basketball Association (NBA), Major League Baseball (MLB), and the National Hockey League (NHL) and the National Football League (NFL).
 - a) Find the mean and median salaries for the entire group of 10 athletes.
 - b) Is the median salary for the entire group of 10 athletes a realistic measure of the "average" salary of highest paid professional athletes? Explain.
 - c) Is the mean salary for the entire group of 10 athletes a realistic measure of the "average" salary of highest paid professional athletes? Explain.
- 24. An instructor grades on a curve (normal distribution) and your grade for each test is determined by the following where S = your score.

A-grade: $S \ge \mu + 2\sigma$ B-grade: $\mu + \sigma \le S < \mu + 2\sigma$ C-grade: $\mu - \sigma \le S < \mu + \sigma$ D-grade: $\mu - 2\sigma \le S < \mu - \sigma$ F-grade: $S < \mu - 2\sigma$

If on a particular test, She states that the average on the test was $\mu = 66$, the standard deviation was $\sigma = 15$. What is the range of scores for each grade? If you got an 81%, what grade did you get on that test?

25. Using the grading scheme in the above problem, what is the range of scores for each letter-grade on a test if the average was $\mu = 75$, the standard deviation was $\sigma = 6$? If you got an 81%, what grade did you get on that test?

Section 4.5 Polls and the Margin of Error

One of the most visible uses of statistics and the normal curve can be found in public opinion surveys. In these polls, a relatively small segment of the population is asked about a certain issue, and the responses are used to predict the views of the entire population. In this section, we will examine how the results of these polls can be interpreted and the accuracy of the predictions determined.

Inferential Statistics and Confidence Intervals

Inferential statistics are the statements about a population that are derived from information about a small segment of a population. In other words, inferential statistics are exactly the types of information that can be derived from a poll or survey. A particular type of inferential statistic is the confidence interval. Before defining this term, consider the following example.

Example 1

Two thousand registered voters have been asked to respond to a questionnaire about whether the state's members of the House of Representatives should be subject to term limitations of eight years. The results of the survey show that 1120 of the 2000 people think that the Representatives should be limited to a maximum of eight years in office. What does this survey say about the entire voting population of the state?

Solution: Since 1120 of the 2000 voters endorse term limitations, it is appropriate to say that approximately $1120 \div 2000 = 0.56$ or 56% of voters endorse term limitations. However, we cannot say that 56% of all the voters endorse the proposal. If we took another survey of a different group of 2000 people, we might find that 1160 or 1070 people endorse term limitations.

Now that we recognize that any survey will produce only an estimate of an actual value, the next step is to determine what factors contribute to the difference between the estimate and an actual value. One factor that will affect the accuracy of the estimate is the size of the sample. If the survey asks a large number of people to respond to a question, the results of the survey should be more accurate than if the survey asks only a small number of people. With an increase in the size of the sample comes an increase in the accuracy of the estimate and an increase in the confidence we have in the validity of our estimate.

An additional consideration comes from the theory of statistics. Suppose that we conduct the poll described in Example 1 one hundred times, using a different sample of the population in each poll. Each poll will have a resulting estimate of the percentage of the people who are in favor of term limitations. By drawing a graph of these proportions, it can be shown that the percentages are distributed in the shape of a normal curve. As was true with the situations described in Section 5.4, the exact shape and position of the normal distribution will depend on the mean and the standard deviation of the sample percentages.

The above discussion leads to two important concepts in statistics, confidence intervals and confidence levels. A **confidence interval** is an interval centered around the estimate gener-

ated by the survey. For instance, in Example 1, the results of the survey stated that 56% of the voters endorsed term limitations. If we allow a margin of error of four percentage points, we anticipate that the true percentage in favor of term limitations could be as low as 52% (56 – 4) and as high as 60% (56 + 4). A confidence interval for this situation would then be 52% to 60%.

A **confidence level** is a statement of the probability that the actual percentage being studied is contained within the confidence interval. With all other factors remaining equal, the higher the confidence level, the larger the confidence interval. With these considerations in mind, it can be shown that the formula for the **margin of error** in a poll or survey is given as follows.

$$M = \frac{z}{2\sqrt{n}} \quad \text{where} \quad \begin{cases} M = \text{ margin of error} \\ z = \text{ value determined by confidence level} \\ n = \text{ number of people surveyed} \end{cases}$$

Confidence Level	<i>z</i> -value
90%	1.645
95%	1.96
98%	2.327
99%	2.575

Margin of error and surveys

Now that we have completed the background material, we can apply the margin of error formula to a variety of situations.

Example 2

A 2004 study found that 75% of Americans had Internet access. If the survey was conducted with a sample size of 1011 adults and had a 95% confidence level, what is the margin of error in the survey?

Solution: Since we have a 95% confidence level, z = 1.96. Using this *z*-value and n = 1011, we have the following.

$$M = \frac{1.96}{2\sqrt{1011}} = 0.03$$

Thus, the margin of error is 3%. This means that the percentage of people who had Internet access is somewhere between 75 - 3 = 72% and 75 + 3 = 78%.

Example 3

Suppose that the pollster in Example 2 wants to decrease the margin of error from 3% to 1% while maintaining the same level of confidence. How large a sample should the pollster use?

Solution: Using M = 0.01 and z = 1.96, we have the following.

$$0.01 = \frac{1.96}{2\sqrt{n}}$$
$$0.02\sqrt{n} = 1.96$$
$$\sqrt{n} = 98$$
$$n = 98^{2}$$
$$n = 9604$$

Notice that the reduction in the margin of error requires a very large increase in the sample size. This is an important consideration for pollsters. While it is desirable to decrease the margin of error, the resulting increase in the sample size could cause a large increase in the cost of the poll.

§4.5 Explain \rightarrow Apply \rightarrow Explore

Explain

- 1. What is a confidence interval?
- 2. What is a confidence level?
- 3. What is a margin of error?
- 4. What is a sample size?
- 5. What happens to the margin of error if the sample size increases? Explain.
- 6. What happens to the margin of error if the confidence level increases? Explain.
- 7. A survey estimates that 25% of the viewing audience watched the Rose Bowl. If there is a 5% margin of error, what does this say about the percentage of the viewing audience that watched the Rose Bowl?
- 8. A poll estimates that 63% of the public wants to see an increase in handgun control laws. The margin of error is 4%. What does this say about the percentage of the public that wants to see an increase in the number of handgun control laws?

Apply

- 9. A poll has a result of 63% with a margin of error of 4%. What is the confidence interval for this poll?
- 10. A poll has a result of 52% with a margin of error of 3%. What is the confidence interval for this poll?
- 11. A poll of 750 people is taken and the confidence level is 98%. What is the margin of error?
- 12. A poll of 1200 people is taken, and the confidence level is 95%. What is the margin of error?
- 13. A poll of 850 people has a margin of error of 4%. What is the confidence level for this poll?
- 14. A poll of 1842 people has a margin of error of 3%. What is the confidence level for this poll?
- 15. A customer requests that a polling company produce a survey with a 98% confidence level and a margin of error of 2%. How many people must respond to the survey?

16. A customer requests that a polling company produce a survey with a 99% confidence level and a margin of error of 1%. How many people must respond to the survey?

Explore

- 17 In a poll of 266 smokers, 52% said that secondhand smoke was very harmful. The poll listed the confidence level at 97.74%. If the margin of error was 7%,
 - a) How large a sample would be needed to have a confidence level of 99% while maintaining a 7% margin of error?
 - b) How large a sample would be needed to decrease the margin of error to 2% while maintaining a 99% confidence level?
- 18. In a poll of 1005 adults, 40% said that the economic conditions in their local economy were getting better. The poll listed the confidence level at 94.26%. If the margin of error was 3%,
 - a) How large a sample would be needed to have a confidence level of 98% while maintaining a 3% margin of error?
 - b) How large a sample would be needed to decrease the margin of error to 2% while maintaining a 98% confidence level?
- 19. In a poll of 506 adults, 74% said that they would be willing to pay \$100 more each year in higher prices so that industry could reduce air pollution. If the confidence level was 95%, what was the confidence interval for the survey.
- 20. In a Gallup poll of adult men, 49% said that they were trying to lose weight. If the margin of error is 3% and the confidence level is 95%, how many men were surveyed?