Binomial Distributions and Statistical Inference

**LESSONS**

1. **Binomial Distributions**
   - Develop the binomial probability formula and construct binomial distributions and compute their expected value. Compute and interpret a P-value for the proportion of successes in a sample, deciding whether the result is statistically significant or can reasonably be attributed to chance alone.

2. **Sample Surveys**
   - Evaluate and design surveys that satisfy the characteristics of a trustworthy sample survey. Distinguish between random sampling and stratified random sampling. Identify sample selection bias and response bias.

3. **Margin of Error: From Sample to Population**
   - Distinguish between point and interval estimates of a parameter. Observe variability in sampling (sampling error) in approximate sampling distributions. Compute and interpret a margin of error and a 95% confidence interval for a proportion. Understand the meaning of 95% confidence.

Public opinion polls gather information only from a relatively small sample of the population. Nevertheless, they can estimate proportions, such as the proportion of voters who approve of the job the president is doing, with surprising precision.

Through work on the investigations of this unit, you will learn how polls are conducted and analyzed. You will learn what characteristics make a public opinion poll trustworthy. Finally, you will construct and use binomial distributions to understand how a poll can measure public opinion to within a specified margin of error.

These key ideas will be developed through your work in the following three lessons.
The ethnicity, gender, age, and other demographic characteristics of juries have been of great interest in some trials in the United States. When the composition of a jury does not reflect the demographic characteristics of the surrounding community, doubts about fairness of the jury selection process and legal challenges can arise.

Although juries are not selected solely by chance, comparing the actual jury to the composition of juries that would occur if jurors were selected at random can tell lawyers whether there are grounds to investigate the fairness of the jury selection process.

An historic case concerning jury selection, Avery v. Georgia, was brought to the U.S. Supreme Court in 1953. A jury in Fulton County, Georgia had convicted Avery, an African-American, of a serious felony. There were no African-Americans on the jury. At the time, there were 165,814 African-Americans in the Fulton County population of 691,797. The list of 21,624 potential jurors had 1,115 African-Americans. A jury pool of 60 people was selected, supposedly at random, from the list of potential jurors. (However, the names of black and white jurors had been written on different colored slips of paper.) This jury pool, from which the 12 actual jurors were selected, contained no African-Americans. (Source: caselaw.lp.findlaw.com/cgi-bin/getcase.pl?court=US&vol=345&invol=559)
THINK ABOUT THIS SITUATION

Think about the demographics of Fulton County, Georgia and of the jury selected for the trial of James Avery.

a. If 12 jurors were selected at random from the people in Fulton County, how can you compute or estimate the probability that there would be no African-Americans on the jury? Generate as many methods as you can.

b. Do you think that having only 1,115 African-Americans on the list of potential jurors reasonably can be attributed to chance alone or should the lawyers look for another explanation? What strategies could you use to support your choice?

c. The jury pool of 60 people was selected from the list of 21,624 potential jurors. Can getting no African-Americans in a jury pool selected from these potential jurors reasonably be attributed to chance alone or should the lawyers look for another explanation? What strategies could you use to support your choice?

d. The U.S. Supreme Court overturned Avery’s conviction. Describe the statistical evidence that you think might have been used by Avery’s lawyers.

In this lesson, you will construct the type of probability distribution that will enable you to analyze situations such as the selection of jurors for Avery’s trial.

INVESTIGATION 1

Rules of Probability and Binomial Situations

Many situations involving probability are called binomial because they have two possible outcomes, success and failure. In a binomial situation, you are interested in counting the number of successes that occur in a fixed number of identical, independent trials. You flip a coin 5 times and count the number of heads. You roll a pair of dice 12 times and count the number of times you get a sum of 7. You pick 50 students at random from your school and count the number who plan to go to a community college.

As you work on the problems in this investigation, look for an answer to this question:

How can rules of probability help you analyze a binomial situation?
This sample space shows all possible outcomes of the roll of a pair of dice, one red and one green. There are six ways the red die can land and six ways the green die can land, so there are 36 possible outcomes. These 36 possible outcomes are equally likely to occur.

<table>
<thead>
<tr>
<th>Number on Green Die</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,1</td>
<td>1,2</td>
<td>1,3</td>
<td>1,4</td>
<td>1,5</td>
<td>1,6</td>
</tr>
<tr>
<td>2</td>
<td>2,1</td>
<td>2,2</td>
<td>2,3</td>
<td>2,4</td>
<td>2,5</td>
<td>2,6</td>
</tr>
<tr>
<td>3</td>
<td>3,1</td>
<td>3,2</td>
<td>3,3</td>
<td>3,4</td>
<td>3,5</td>
<td>3,6</td>
</tr>
<tr>
<td>4</td>
<td>4,1</td>
<td>4,2</td>
<td>4,3</td>
<td>4,4</td>
<td>4,5</td>
<td>4,6</td>
</tr>
<tr>
<td>5</td>
<td>5,1</td>
<td>5,2</td>
<td>5,3</td>
<td>5,4</td>
<td>5,5</td>
<td>5,6</td>
</tr>
<tr>
<td>6</td>
<td>6,1</td>
<td>6,2</td>
<td>6,3</td>
<td>6,4</td>
<td>6,5</td>
<td>6,6</td>
</tr>
</tbody>
</table>

a. Suppose that you roll a pair of dice, one red and one green. Use the sample space to find the probability that you get doubles (both dice show the same number). What is the probability that you do not get doubles?

b. Now suppose that you roll two dice that are the same color. Why should the sample space remain the same as that above? What is the probability that you get a sum of 7? What is the probability that you do not?

c. What is the probability that you get a sum of 6? What is the probability that you do not?

d. Complete this table that shows the probability distribution for the sum of the two dice. Is it best to leave your answers as unreduced fractions, reduced fractions, or decimals?

<table>
<thead>
<tr>
<th>Sum</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

e. Which sum is the most likely? Which sums are the least likely?
2 Now suppose you roll a pair of dice two times. Because the results of successive rolls of a pair of dice are independent events, you can compute probabilities using the **Multiplication Rule for Independent Events**:

If \( A \) and \( B \) are independent events, the probability, \( P(A \text{ and } B) \), that event \( A \) occurs and event \( B \) occurs is:

\[
P(A \text{ and } B) = P(A) \cdot P(B)
\]

\( P(A) \) denotes the probability that \( A \) occurs on the first trial and \( P(B) \) the probability that \( B \) occurs on the second trial.

Formally, two events, \( A \) and \( B \), are **independent** if and only if the rule above holds. An equivalent definition is that events \( A \) and \( B \) are independent if the occurrence of one of them does not change the probability that the other event occurs.

a. Show how to use this rule to compute the probability that you get a sum of 7 on both rolls of a pair of dice.

b. Show how to use this rule to compute the probability that you get a sum of 6 both times.

c. Show how to use this rule to compute the probability that you get a sum of 7 on the first roll of the dice, but not on the second roll.

d. Compute the probability that you do not get a sum of 6 on the first roll of the dice, but do on the second roll.

3 The Multiplication Rule for Independent Events can be extended to situations where there are more than two independent trials:

In \( n \) independent trials, the probability \( P(A \text{ and } B \text{ and } C \text{ and } \ldots) \) of getting \( A \) on the first trial, \( B \) on the second trial, \( C \) on the third trial, and so on is

\[
P(A \text{ and } B \text{ and } C \text{ and } \ldots) = P(A) \cdot P(B) \cdot P(C) \cdot \ldots
\]

a. Use this rule to compute the probability that if you roll a pair of dice five times, you get a sum of 7 on each roll. Then, find the probability that you get a sum of 7 on the first roll, but not on the remaining four rolls.

b. Use this rule to compute the probability that if you flip a coin ten times, you get a head on each flip. What is the probability that you do not get 10 heads?

c. Suppose that you roll a pair of dice six times. Find the probability that you never get doubles.

d. Suppose you roll a pair of dice six times. Recall that the sum of the probabilities of all possible outcomes must add up to 1. Use this fact and the results of Part c to find the probability that you get doubles at least once.
The New York Lottery has a game called Quick/Draw. In one version of the game, you pick a number from 1, 2, 3, …, 78, 79, 80. The lottery computer then randomly selects 20 of the 80 numbers. If the number you picked is included in the 20 randomly selected numbers, you win.

a. What is the probability you win if you play the game once?

b. If you play the game twice, what is the probability that you win both times?

c. If you play this game 10 times, what is the probability that you win all 10 times?

d. If you play this game 10 times, what is the probability that you lose all 10 times?

e. Use the result of Part d to find the probability that if you play this game 10 times, you win at least once.

5 Suppose that you conduct $n$ independent binomial trials. The probability of a success $p$ is the same on each trial.

a. What is the probability that all $n$ trials are successes?

b. What is the probability of a failure on a trial? What is the probability that all $n$ trials are failures?

c. What is the probability that you get at least one success?

6 Another rule of probability that is helpful in binomial situations is called the Addition Rule for Mutually Exclusive Events.

a. Refer to the sample space in Problem 1. If you roll two dice, what is the probability that you get a sum of 2? What is the probability that you get a sum of 5? What is the probability that you get a sum of 2 or a sum of 5?

b. In the last question of Part a, to find the probability of getting a sum of 2 or a sum of 5 when rolling a pair of dice once, you may have used the Addition Rule for Mutually Exclusive Events:

If event $A$ and event $B$ are mutually exclusive, $P(A \text{ or } B) = P(A) + P(B)$.

Mutually exclusive means that the two events cannot happen on the same trial. For example, you cannot get both a sum of 2 and a sum of 5 when you roll a pair of dice just once. Sometimes the term disjoint is used instead of mutually exclusive.

Use the rule to find the probability that if you roll two dice, you get doubles or a sum of 7.

c. Can you use the rule in Part b to find the probability that, if you roll two dice, you get doubles or a sum of 8? If so, do it. If not, explain why not.

d. Can you use the rule in Part b to find the probability that if you flip a coin four times, you get either all heads or all tails? If so, do it. If not, explain why not.
In problems involving rolling dice and flipping coins, the trials are identical. For example, when playing a game using dice, the probability of rolling a sum of 7 does not change from one turn to the next. However, when playing a game using a deck of cards, once a card has been drawn, it usually is removed from the deck. This is called sampling without replacement. Consequently, the probability of getting, say, a heart on the next draw depends on what cards have been drawn before.

7 In this problem, you will compute a probability two different ways. Suppose a town has a pool of 3,000 eligible voters. Of these, 1,800 are college graduates. A four-person focus group is selected at random from the eligible voters and the number of college graduates is counted.

a. What percent of the population is the sample size?

b. What is the probability that the first member of the focus group selected is a college graduate?

c. If the first member selected was a college graduate and is removed from the pool, what is the probability that the second member selected is a college graduate? Continuing in this way, compute the probability that all four members of the focus group are college graduates.

d. Now suppose that each name is replaced in the pool after it is drawn. (This is called sampling with replacement.) Compute the probability that all four members of the focus group are college graduates.

e. Is there much difference in the probability from Part c and the probability from Part d that all four members are college graduates? Which was easier to calculate?

Your work in Problem 7 suggests the following guideline used by statisticians to tell whether the probability changes significantly from trial to trial or whether the change is so small you can ignore the fact that you are sampling without replacement when computing probabilities.

Sample Size Guideline
If the size of a random sample is less than 10 percent of the size of the population from which it is taken, then, without much loss of accuracy, you can ignore the fact that the sampling is without replacement when computing a probability.
Test this guideline using the following situation. Now suppose that there are only 30 eligible voters in a town and that 18 are college graduates. A four-person focus group is selected at random from the eligible voters. Repeat Parts a–e of Problem 7 for this new situation. Does the sample size guideline appear to be reasonable?

### SUMMARIZE THE MATHEMATICS

In this investigation, you used rules of probability to compute probabilities in binomial situations.

a. What is the Multiplication Rule for Independent Events? When can it be used?

b. If you conduct \( n \) (independent) binomial trials and the probability of a success on each trial is \( p \), what is the probability that all \( n \) trials are a success? What is the probability that all \( n \) trials are a failure?

c. In a binomial situation, how can you find the probability of getting at least one success?

d. What is the Addition Rule for Mutually Exclusive (Disjoint) Events? When can it be used?

e. When computing probabilities, when can you ignore the fact that the trials are not independent?

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

### CHECK YOUR UNDERSTANDING

Sixty-one percent of 18-year-olds in the United States have a driver’s license. (Source: Science Daily, August 3, 2012, www.sciencedaily.com/releases/2012/08/120803082905.htm) You will select five different 18-year-olds at random from the U.S. and count the number with a driver’s license.

a. Is this a case of sampling with or without replacement?

b. If your first 18-year-old has a driver’s license, is the probability that the second 18-year-old has a driver’s license a bit smaller than 61%, exactly equal to 61%, or a bit larger than 61%? Verify your answer using the fact that there are about 4,389,000 18-year-olds in the U.S.

c. According to the sample size guideline, can you compute the probability that all five 18-year-olds have a driver’s license using \( p = 0.61 \) for each trial, without much loss of accuracy?

d. What is the probability that the first 18-year-old has a driver’s license, but the other four do not?

e. What is the probability all five have a driver’s license? What is the probability that none of the five 18-year-olds have a driver’s license?

f. Find the probability that at least one of the 18-year-olds has a driver’s license.
The Binomial Probability Formula

In a binomial situation, you have a population made up of “successes” and “failures,” where the proportion of the population that are successes is denoted by $p$. In this investigation you will use counting methods to compute binomial probabilities.

As you work on the problems in this investigation, look for answers to this question:

*How can you find the probability of getting a specified number of successes $x$ in a binomial situation with $n$ trials and probability $p$ of success on each trial?*

1. Having a baby and noting the sex is a binomial situation. About 51% of all babies born in the United States are boys. Suppose that a couple is going to have four children.

   a. Compute the probability that all four children will be boys.

   b. Name the rule that was used for the computation in Part a. What condition needs to be in place in order to use this rule?

   c. Compute the probability that all four children will be girls.

2. Now consider the following reasoning used by Elena to compute the other possibilities for a family of four children.

   To compute the probability of getting one boy and three girls, she calculated
   \[
   (0.51)(0.49)(0.49)(0.49) \approx 0.060.
   \]

   To compute the probability of getting two boys and two girls, she calculated
   \[
   (0.51)(0.51)(0.49)(0.49) \approx 0.062.
   \]

   To compute the probability of getting three boys and one girl, she calculated
   \[
   (0.51)(0.51)(0.51)(0.49) \approx 0.065.
   \]
a. Place Elena’s probabilities and your probabilities from Problem 1 Parts a and c in a copy of the table below.

<table>
<thead>
<tr>
<th>Number of Boys in a Family of Four Children</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

b. Why can’t the probabilities in the table all be correct?

c. What is wrong with Elena’s reasoning?

3 You can use careful counting of the possible sequences of births to complete the table in Problem 2 correctly.

a. Imagine all possible families with three boys and one girl.
   i. List all of the sequences of births that would result in a family with three boys and one girl.
   ii. For each of the possible birth sequences, compute the probability that it will occur. What do you notice?
   iii. Find the probability that a family of four children will have three boys and one girl (in any order). What probability rule did you use and why?

b. Imagine all possible families with two boys and two girls.
   i. List all of the birth sequences that would result in a family with two boys and two girls.
   ii. For each of the birth sequences, compute the probability it will occur. What do you notice?
   iii. What is the probability that a family of four children will have two boys and two girls?

c. What is the probability that a family of four children will have one boy and three girls?

d. Using your results from Parts a, b, and c, place the correct probabilities in a copy of the table in Problem 2. Check to be sure that the probabilities in the table add up to 1 (subject to round-off error).
4 In Problem 3, you found the number of possible birth sequences for families of four children by listing them. If you use the idea of combinations from Unit 3, you do not have to list all possibilities.

a. Compute \( C(4, 3) \). Why does \( C(4, 3) \) give you the number of ways that a family of four children could have three boys and one girl?

b. Use combinations to find the number of ways that a family of four children could have two boys and two girls.

c. Use combinations to find the number of ways that a family of four children could have one boy and three girls.

d. How many ways could a family of \( n \) children have exactly \( x \) boys?

5 Suppose a softball player has a batting average of .400. In the next game, she expects to be at-bat five times. One model that statisticians have investigated is whether a player’s at-bats are independent. Independence would mean that the player has a 0.4 chance of making a hit each time she comes up to bat, no matter what has happened in previous at-bats. In this problem, assume that at-bats are independent; that is, she does not tend to have streaks or slumps that require an explanation other than chance.

a. According to the Multiplication Principle of Counting, how many different sequences of hits and outs are there for five at-bats?

b. Explain why the entries for 4 hits in columns 2 and 3 of the following probability distribution table are correct.

<table>
<thead>
<tr>
<th>Number of Hits</th>
<th>Number of Possible Sequences</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( C(5, 0) = 1 )</td>
<td>1 ((0.4)^0(0.6)^5 \approx 0.078 )</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( C(5, 4) = 5 )</td>
<td>5 ((0.4)^4(0.6) \approx 0.077 )</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c. Using the same format as in the rows already entered, complete a copy of the table above.
d. Check your results to be sure you have the right total number of possible sequences and that the probabilities add up to 1.

e. Describe the patterns you see in the table.

f. Make a histogram of the probability distribution similar to the one begun below. Then describe its shape and center.

![Histogram of Probability Distribution]

g. Using the graph, what number of hits has the largest probability?

6 In the United States, a person has a right to control inventions, literary and artistic works, and other ideas that they have created. This is called intellectual property rights. About 30% of adults are college graduates. Suppose that this is true in a jury pool where, in a case about intellectual property rights, the jury of 12 people contains only two college graduates. (Source: U.S. Census Bureau, 2011 American Community Survey)

Catherine and Houston decided to build a spreadsheet for computing the probabilities for this situation. The first rows of their spreadsheet are shown below.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of College Graduates</td>
<td>Number of Possible Sequences</td>
<td>Probability of Each Particular Sequence</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>=COMBIN(12,A2)</td>
<td>=(0.3)^(A2)*(1-0.3)^(12-A2)</td>
</tr>
<tr>
<td>3</td>
<td>=A2+1</td>
<td>=COMBIN(12,A3)</td>
<td>=(0.3)^(A3)*(1-0.3)^(12-A3)</td>
</tr>
</tbody>
</table>

a. Discuss how the entries in rows 2 and 3 will compute the probabilities of 0 college graduates and 1 college graduate.

b. What should be the entries of row 4 of this spreadsheet? Compare your entries with those of your classmates. Resolve any differences.
c. Complete the spreadsheet and display a frequency distribution chart as shown below.

![Frequency Distribution Chart]

d. Describe the center and spread of the distribution.

e. What is the probability of getting two or fewer college graduates on a jury that is selected at random?

f. How can the probability in Part e be roughly estimated by comparing the sum of the areas represented by the first three bars and the total area represented by the bars?

g. Can you reasonably attribute the composition of the jury in the intellectual property case to chance alone or should the lawyers look for some other explanation? Explain your reasoning.

7 Refer to your work in Problems 5 and 6. Write the formula you have discovered for the probability $P(x)$ of getting exactly $x$ successes in a binomial situation with $n$ trials and probability of success $p$ on each trial. Be sure to define any symbols you use.


a. Use your formula from Problem 7 to find the probability that exactly four of them have ADHD.

b. Find the probability that at least one of these children has ADHD.

c. Produce a graph of this probability distribution.

d. What is the probability that four or fewer children have ADHD?

9 The mean of a probability distribution, also called the expected value, is the value you would get, on average, in the long run if you repeated the series of trials again and again. The expected value $\mu$ of a binomial distribution with probability of success $p$ and $n$ trials may be found using the formula $\mu = np$.

a. Refer to Problem 1 of Investigation 1. You will roll a pair of dice 60 times. What is the expected number of times you will get a sum of 7?
b. Refer to Problem 6. What is the expected number of college graduates on a randomly selected jury of 12 people?

c. Refer to Problem 4 of Investigation 1. It costs $1 to play this lottery game. If you win, you are paid $2. What is your expected net gain (or loss) if you play this game 20 times?

d. Expected value can be used to help price insurance. For example, suppose that a company insures people against being struck by lightning and expects to sell 3,100,000 policies. The probability of being struck by lightning in a year is \( \frac{1}{775,000} \). (Source: www.lightningsafety.noaa.gov/odds.htm) If an insured person is struck by lightning, the company would pay them $1,000,000. What is the expected number of insured people who will be struck by lightning? What is the expected total payout to them? What should the company charge each insured person per year (called a premium) in order to expect to break even?

**SUMMARIZE THE MATHEMATICS**

In this investigation, you developed a formula for computing binomial probabilities.

a. Describe your formula for computing binomial probabilities and explain why it works.

b. In what situations can you use your formula to compute binomial probabilities? Explain your reasoning.

c. What is the expected number of successes in a binomial situation with probability of success \( p \) and \( n \) trials? Does the expected number of successes have to be a whole number? Explain your reasoning. Give an example to illustrate this computation.

*Be prepared to explain your ideas and reasoning to the class.*

**CHECK YOUR UNDERSTANDING**

In the Avery v. Georgia case from the Think About This Situation on page 341, the list of 21,624 potential jurors in Fulton County had 1,115 African-Americans. Suppose that a 12-member jury is randomly selected from the list of potential jurors.

a. What is the probability that the first juror selected is African-American? Is it reasonable to use this same probability for each trial? Explain why or why not.

b. What is the expected number of African-Americans who would be on a jury of 12 people?
c. Complete this probability distribution table for the number of African-Americans on a jury that really was randomly selected from the list of potential jurors.

<table>
<thead>
<tr>
<th>Number of African-Americans, x</th>
<th>Probability, P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3456</td>
</tr>
<tr>
<td>2</td>
<td>0.1033</td>
</tr>
<tr>
<td>3</td>
<td>0.0187</td>
</tr>
<tr>
<td>4</td>
<td>0.0023</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

d. Describe the shape of the probability distribution below. What is the most likely number of African-Americans on a randomly selected jury?

e. Can you reasonably attribute the fact that there were no African-Americans on a jury selected from this list of potential jurors to chance alone or should Avery’s lawyers look for some other explanation? Explain your reasoning.
Statistical Significance

Astrologers claim that natal charts, a type of horoscope based on where and when a person was born, can be used to predict personality.

<table>
<thead>
<tr>
<th>Sun</th>
<th>Moon</th>
<th>Mercury</th>
<th>Venus</th>
<th>Mars</th>
<th>Jupiter</th>
<th>Saturn</th>
<th>Uranus</th>
<th>Neptune</th>
<th>Pluto</th>
</tr>
</thead>
<tbody>
<tr>
<td>27.23</td>
<td>8.21</td>
<td>11.23</td>
<td>8.20</td>
<td>9.15</td>
<td>2.24</td>
<td>4.06</td>
<td>21.48</td>
<td>4.59</td>
<td>4.02</td>
</tr>
</tbody>
</table>

To test this claim, astrologers prepared natal charts for 83 volunteer subjects and wrote a description of the person's personality based only on the chart. Each subject then was given three descriptions (their own and two randomly chosen descriptions that were made for other people). The subject was asked to pick out the one that most correctly described them. Twenty-eight out of 83 (about 34%) selected their own description. (Source: Shawn Carlson, “A Double-Blind Test of Astrology,” Nature, Vol. 318, pages 419–425, December 5, 1985.) If natal charts are complete nonsense, you would expect about 33% of the volunteers to pick their own description just by chance. Getting 34% who pick their own description certainly is not convincing evidence that natal charts can be used to predict personality. So, we say that the result from the experiment is not statistically significant. It reasonably can be attributed to chance alone.

As you work on the problems in this investigation, look for answers to this question:

How can you use technology to find out whether the number of successes in a binomial situation is statistically significant?

1 You often will find it best to use technology to compute binomial probabilities. Explore the binomial probability functions of your calculator or TCMS-Tools as you answer the following questions.

To use your calculator to compute, for example, the binomial probabilities associated with selecting 12 jurors at random from a large population in which 30% of eligible jurors are college graduates, enter the DISTR menu and select \( \text{binompdf} \). (The initials pdf stand for probability density function.) Then enter the number of trials \( n \), the probability of a success \( p \), and the number of successes \( x \).
To compute this value in TCMS-Tools, open the CAS, select “Auto Numeric” from the Options menu, and type `binompdf(12, 0.3, 3)` in the Command window as shown below.

![Image of CAS screen with binompdf function]

**a.** Use the `binompdf(n, p, x)` function of your preferred technology tool to answer these questions. If 30% of people in a large city are college graduates and you select 12 people at random:

i. what is the probability that exactly three are college graduates?

ii. what is the probability that three or fewer are college graduates?

iii. Use your probability from part ii to find the probability that four or more are college graduates.

**b.** Use a technology-based `binompdf(n, p, x)` function to answer these questions. If 18% of the seniors in a large high school have their own car and you pick 20 seniors at random:

i. what is the probability that exactly 2 have their own car?

ii. what is the probability that 2 or fewer have their own car?

iii. what is the probability that 3 or more have their own car?

**c.** Use a technology-based `binomcdf(n, p, x)` function to evaluate `binomcdf(12, 0.3, 3)`.

(The initials `cdf` stand for *cumulative distribution function.*) Compare this answer to your answers in Part a.

i. Explain what the `binomcdf` function does.

ii. Use the `binomcdf` function to answer the second and third questions in Part b.

**d.** Find the probability of getting 45% or fewer heads if you flip a coin 400 times. Why would you not want to do this without technology capable of computing binomial probabilities?

**e.** Suppose you roll a pair of dice 100 times. Find the probability that you will roll doubles on at least 20% of the rolls.
2 Suppose your family has been playing a board game where the player spins a plastic spinner on each turn. The spinner is divided into five equal sections, and one of the sections says, “Go home and start again.” You have watched 50 different spins and the player had to “Go home and start again” on 14 of them. You are beginning to think that the spinner is unbalanced so that it is too likely to land on “Go home and start again.”

a. If the spinner is balanced, what is the expected number of “Go home and start again” in 50 spins?

b. If the spinner is balanced fairly, what is the probability that the player will have to “Go home and start again” on 14 or more out of 50 spins? Answer this problem two ways.

i. Estimate the probability from this graph of the binomial distribution constructed using $n = 50$ and $p = 0.2$.

ii. Compute the probability using the \texttt{binomcdf} function. Is it unusual to get “Go home and start again” 14 times out of 50 spins if the spinner is balanced?

The probability you computed in Part b is called a $P$-value. To compute a $P$-value, you start with a value of a population proportion $p$ against which you will compare the actual result of $x$ successes out of $n$ trials. Then, using that value of $p$, you compute the probability of getting, just by chance, $x$ or even more successes (or, in some scenarios, $x$ or even fewer successes) in $n$ trials. When the $P$-value is less than 0.05, standard statistical practice is to call the result statistically significant. You want to look carefully at any actual result that is statistically significant (has low probability of happening just by chance) to see if you can understand why such an unusual event happened.

c. Which of the following is the best conclusion about the spinner?

I You do not have statistically significant evidence that the spinner is unbalanced.

II You do have statistically significant evidence that the spinner is unbalanced.

III The spinner must be balanced.

IV The spinner must be unbalanced.

3 In many sports, the home team is more likely to win a game than is the visiting team. This is called the “home-field advantage.” For example, in the 2012 Major League Baseball season, 2,430 games were played. The home team won 1,295 of them. [Source: www.baseball-reference.com/games/situational.shtml]
a. If there is no home-field advantage, what proportion of the time do you expect that the team playing at home will win? What was the actual proportion in 2012?

b. What is the P-value for this situation? That is, if there is no home-field advantage, what is the probability that the home team will win 1,295 or more of 2,430 games? Answer this problem two ways.

i. Estimate the probability from the graph of the binomial distribution. This distribution was constructed using \( n = 2,430 \) and \( p = 0.5 \).

ii. Compute the probability using the binomcdf function.

c. Is the number of wins by the home team statistically significant? If so, make several conjectures about what the explanation might be.

4 For a project in statistics class, Miguel individually showed 75 adults two different horoscopes taken from yesterday’s newspaper. Each adult picked the horoscope that best described what happened to him or her yesterday. The adult did not know it, but one of the horoscopes was the one in the newspaper for his or her sign of the Zodiac and the other was for a different sign. Of the 75 adults, 39 picked the horoscope for their sign of the Zodiac.

a. If horoscopes have no basis in reality, should the P-value for this situation be larger or smaller than 0.05? Why?

b. What is the probability that an adult will pick his or her own horoscope? What is the probability that 39 or more out of 75 adults will pick their own horoscope?

c. Which of the probabilities in Part b is the P-value for this situation? Is Miguel’s result statistically significant?

d. The distribution below illustrates Miguel’s situation. Describe how this distribution was constructed. Then describe how to tell from the graph alone (even without a scale on the vertical axis) whether Miguel’s result is statistically significant.

 e. What should Miguel conclude?
5 An important 1977 U.S. Supreme Court case, Castaneda v. Partida, was explicitly decided on statistical evidence. A grand jury is responsible for deciding whether a person will have to stand trial. Of the 870 persons who were summoned to serve as grand jurors in Hidalgo County, Texas over an 11-year period, 339, or 39%, were Spanish surnamed. Census figures showed that 79.1% of the county’s population had Spanish surnames. Castenada was convicted of a crime in Hidalgo County, Texas and appealed. (Source: caselaw.lp.findlaw.com/scripts/getcase.pl?court=US&vol=430&invol=482) While grand jurors were not selected by chance in those days (as they typically are now), comparing the actual jurors to what might happen if jurors were selected randomly can help us decide whether there is any evidence of possible unfairness in the selection process.

a. Compute the P-value for this situation. That is, suppose that 870 people are selected at random from the population of Hidalgo County. Use the binomcdf function to find the probability that 339 or even fewer have Spanish surnames.

b. Is the result from the Hidalgo County sample statistically significant? That is, can getting 339 or even fewer who are Spanish surnamed reasonably be attributed to chance alone or should the court look for another explanation?

c. Think of an explanation, other than discrimination, that could have resulted in such a small percentage of grand jurors who had Spanish surnames. What data could be collected to provide evidence for or against that explanation?

6 In 2014, eighty men and twenty women served as United States senators. About 51% of the adult population of the U.S. are women.

a. Suppose that U.S. senators are selected by a process that is totally without regard to sex. The graph below shows the binomial distribution for this situation. What is $n$? What is $p$?

b. From the graph alone, decide whether the number of women in the Senate in 2014 was statistically significant. What can you conclude?
It used to be the case that 30% of the pieces of a well-known chocolate candy had a brown coating. Sam thinks that the proportion is smaller now. To test this, he gets a random sample of 185 candies. Only 41 had a brown coating.

a. If you take a random sample of 185 chocolate candies from a population where 30% have a brown coating, what is the probability of getting 41 or fewer that have a brown coating?

b. What should Sam conclude?

SUMMARIZE THE MATHEMATICS

In this investigation, you learned that binomial distributions and the idea of statistical significance are useful not only in examining composition of juries, but also in many other situations.

a. What methods can you use to find binomial probabilities?

b. What is the meaning of a P-value? How do you compute it?

c. How can you tell when a result is statistically significant? What does it mean if a result is statistically significant?

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

CHECK YOUR UNDERSTANDING

In the Think About This Situation on page 341, you read about the Avery v. Georgia case.

a. There were 165,814 African-Americans in the Fulton County population of 691,797. What proportion is this?

b. The list of 21,624 potential jurors in the county had 1,115 African-Americans. What proportion is this?

c. What is the P-value? That is, what is the probability of getting 1,115 or even fewer African-Americans if the 21,624 potential jurors were selected at random from the population of Fulton County?

d. Can the composition of the list of potential jurors in Fulton County reasonably be attributed to chance alone? Why or why not?
ON YOUR OWN

APPLICATIONS

1 Cystic fibrosis (CF) is a serious genetic disease. Approximately one in 31 Americans is an unknowing symptomless carrier of the defective gene. A person can be born with CF only if both parents are carriers of the defective gene. In such cases, the child has a 25% chance of being born with CF. (Source: www.cff.org/AboutCF/Testing/) Suppose a couple plans to have three children, and both parents are symptomless carriers of the disease.

a. What is the probability that all of the children are born with CF?

b. What is the probability that none of the children is born with CF?

c. What is the probability that at least one of the children is born with CF?

d. What is the probability that the oldest child is born with CF but the others are not?

e. Suppose a father and a mother are picked at random. What is the probability that they both are symptomless carriers of CF?

2 According to the U.S. Census Bureau, almost half a million people age 5 or older live in Tucson, Arizona. Of these, 33.5% speak a language other than English at home. (Source: quickfacts.census.gov/qfd/states/04/0477000.html) Suppose that you select six people age 5 or older at random from the population of Tucson.

a. What is the probability that all of them speak a language other than English at home?

b. What is the probability that none of them speak a language other than English at home?

c. What is the probability that at least one of them speaks a language other than English at home?

d. What is the probability that the first two people you pick speak a language other than English at home but the last four speak English?

3 California has more residents than any other state. According to the 2010 U.S. census, there are 308,745,538 residents of the United States and 37,253,956 of them live in California. (Source: www.census.gov/2010census/news/releases/operations/cb10-cn93.html) Suppose that you pick five U.S. residents at random. In this task, do not round your answers, but record the entire number given by your calculator.

a. Use the numbers above to find the probability that the first person selected lives in California.

b. If the first person selected lives in California, what is the probability that the second person selected lives in California?
c. Are the events first person selected lives in California and second person selected lives in California independent events? That is, does the result of the first selection change the probability that the second person lives in California?

d. Does the fact that the trials are not independent make much difference in the probabilities in this situation?

4 In the game of Yahtzee®, each player rolls five dice at once. Suppose you are at the end of a game and you need to roll sixes.

a. Find the probability that none of the five dice will show a six.
b. Find the probability that exactly two of the dice will show sixes.
c. Construct the probability distribution table and its graph for the number of sixes.

5 A softball player bats .350. Assume that every time she comes up to bat, the chance she will get a hit is 0.35. In an upcoming series, she will be at bat 10 times.

a. Compute the missing entries in this probability distribution table for the number of hits she will get in the upcoming series.

<table>
<thead>
<tr>
<th>Number of Hits, x</th>
<th>Probability, P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0725</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.2522</td>
</tr>
<tr>
<td>4</td>
<td>0.2377</td>
</tr>
<tr>
<td>5</td>
<td>0.1536</td>
</tr>
<tr>
<td>6</td>
<td>0.0689</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0043</td>
</tr>
<tr>
<td>9</td>
<td>0.0005</td>
</tr>
<tr>
<td>10</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>
b. Suppose the player gets only two hits in the upcoming series of games. What is the probability of getting two or fewer hits just by chance?

c. Some people believe that batters have “slumps” and “streaks.” What do they mean by this? Do these people believe that hits are independent events? Explain.

6 In 1968, during the Vietnam War, Dr. Benjamin Spock (1903–1998) was on trial in Boston for conspiracy to violate the Military Service Act. He was accused of counseling young men on methods of avoiding the draft. It was thought that women would be more sympathetic to Dr. Spock because a larger percentage of women were opposed to the Vietnam War and because many women had used his book about childcare. After a jury selection process of several stages, Dr. Spock ended up with a jury of twelve men, even though women made up more than half of the eligible jurors.

a. Assume that women make up 50% of the eligible jurors and that jurors are selected at random from those eligible. Complete this probability distribution table and graph for the number of women on a Boston jury consisting of 12 members.

<table>
<thead>
<tr>
<th>Number of Women</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0002</td>
</tr>
<tr>
<td>1</td>
<td>0.0029</td>
</tr>
<tr>
<td>2</td>
<td>0.0161</td>
</tr>
<tr>
<td>3</td>
<td>0.0537</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Program: MMH Core Plus Math
Component: TCMS-SE
Vendor: Six Red Marbles
Grade: 12
Final Pass
b. What is the $P$-value for the number of women on Dr. Spock’s jury? That is, what is the probability of getting no women on the jury just by chance? Is this a statistically significant result?

c. If jurors are selected at random, what is the expected number of jurors who would be women? How is the value seen in the graph of the probability distribution?

7 In 1995, O.J. Simpson, an African-American former professional football player, was brought to trial for murder. The large jury pool in Los Angeles was 28% African-American. The final jury, which acquitted Simpson, consisted of nine African-Americans. ([Source: www.law.umkc.edu/faculty/projects/ftrials/Simpson/Jurypage.html])

a. Suppose that jurors were selected at random from the large jury pool without regard to race. The graph below shows the probability distribution for the number of African-Americans on a jury of 12 members. Use the graph to estimate the $P$-value for the actual situation. That is, what is the probability that a jury selected at random would consist of nine or more African-Americans?

b. What is the expected number of African-Americans on a randomly selected jury of 12 people? How is this value seen in the graph?

c. Can the composition of this jury reasonably be attributed to chance alone? Explain.
ON YOUR OWN

8 In 2010, approximately 16% of the population of the United States did not have health insurance. (Source: www.census.gov/newsroom/releases/archives/income_wealth/cb11–157.html) To see whether this percentage has changed, this year a polling organization takes a random sample of 1,500 U.S. residents and counts the number without health insurance.

a. What is \( n \)? Is it reasonable to compute probabilities as if the selections were independent? Explain your reasoning.

b. If the percentage without health insurance remains at 16%, what is the expected number of people in the sample who do not have health insurance?

c. Suppose that 207 people in this year’s random sample of 1,500 U.S. residents do not have health insurance. What percentage is this?

d. What is the \( P \)-value for the situation in Part c? That is, if 16% of the population does not have health insurance, what is the probability that if 1,500 U.S. residents are selected at random that 207 or fewer would have health insurance?

e. Is the difference from the 2010 percentage statistically significant? What can you conclude?

9 In 2010, approximately 15% of the population of the United States lived below the poverty level. (Source: www.census.gov/newsroom/releases/archives/income_wealth/cb11–157.html) To see if the percentage has changed, this year a polling organization takes a random sample of 1,200 U.S. residents.

a. What is \( n \)? Is it reasonable to compute probabilities as if the selections were independent? Explain.

b. If the percentage living below the poverty level remains at 15%, what is the expected number of people in the sample who live below the poverty level?

c. Suppose that 166 people in this year’s random sample of 1,200 residents live below the poverty level. What percentage is this?

d. If the percentage has not changed, what is the probability of getting exactly 166 in the sample who live below the poverty level?

e. If the percentage has not changed, what is the probability of getting 166 or fewer in the sample who live below the poverty level?

f. Which probability, the one in Part d or the one in Part e, is the \( P \)-value?

g. Is the difference from the 2010 percentage statistically significant? What can you conclude?
Look back at the probability distribution table you constructed for the sum of two dice (page 342).

a. Make a graph of the information in the table. Use the horizontal axis for sums and the vertical axis for probabilities.
b. What is the shape of the distribution?
c. What is the mean of the probability distribution? How can the mean be estimated from the histogram?

Explore how to compute probabilities using the Multiplication Rule for Independent Events in cases where the events are not identical binomial trials.

a. You roll a pair of dice two times. Compute the probability that you get a sum of 6 on the first roll and a sum of 7 on the second roll.
b. You roll a pair of dice two times. Compute the probability that you get a sum of 7 on the first roll and a sum of 12 on the second roll.
c. You roll a pair of dice two times. Compute the probability that you get a sum of 7 or a sum of 11 on the first roll and doubles on the second roll.

Sometimes two events are not mutually exclusive (disjoint). For example, suppose that you pick a number from 1 to 10 at random. What is the probability that it is even or prime? If you use the rule in Investigation 1, Problem 6, you would get \( \frac{5}{10} + \frac{4}{10} = \frac{9}{10} \). But only eight numbers—2, 3, 4, 5, 6, 7, 8, and 10—are even or prime, so the correct probability is \( \frac{8}{10} \).

a. Where did the rule from Investigation 1, Problem 6 go wrong?
b. When events \( A \) and \( B \) are not mutually exclusive (disjoint), explain why it is sensible to use the following rule to find \( P(A \text{ or } B) \).

\[
P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)
\]
c. Suppose that 30% of the students in your high school wash their hair in the shower while facing the water, 40% put catsup directly on their fries rather than on the plate, and 15% do both.

If you select one student at random, use the rule to compute the probability that he or she does one or the other (or both). Draw a Venn diagram that illustrates this situation.

d. You roll a pair of dice. Use the rule in Part b to compute the probability that you get a sum of 6 or doubles.
e. Use the rule in Part b to compute the probability that if you draw one card from a deck it is a king or a club.
13 Examine the following tree diagram. Here, \( p \) is the probability of a success and \( q = 1 - p \) is the probability of a failure in a binomial situation.

a. Examine the diagram below. What patterns do you see? What does the third circle in the bottom row represent?

b. Record the next row. What does the third circle in this row represent?

c. Expand \((p + q)^5\).

d. Examine the terms of the expansion. What could they represent?

e. What is the numerical value of \( p + q \)? Of \((p + q)^5\)?

f. What have you just proved?

14 A population of ten items has four “successes” and six “failures.”

SSFFSSFFSS

a. What is the proportion of successes?

b. You can code this binomial population by renaming each success as a 1 and each failure as a 0:

1101000100

Find the mean of these ten numbers.

c. Suppose that you have a population that contains 16 successes and 4 failures.

i. When this population is coded as above, how many 1s will there be? How many 0s?

ii. Find the mean of the coded population.

d. Develop a formula for the mean of a coded population of size \( n \) with \( k \) 1s and \((n - k)\) 0s. Write your formula in a simpler form using the proportion of successes \( p \) in the population.
15 This device is called a binostat. Balls begin at the top, drop through the pegs, and collect in the nine columns at the bottom. As a ball drops from the top of the binostat to a peg where a branch occurs, the ball is equally likely to go right or left.

a. Can the binostat be considered a binomial situation? Explain.
b. Is the distribution of balls in the columns of the binostat shown to the right typical?
c. What do the number labels on the channels represent? Is this pattern of numbers familiar?
d. Suppose the columns at the bottom are numbered 0, 1, 2, … , 8 from left to right.
   i. How many different routes could a ball take to Column 0? What is the probability of a ball falling into Column 0?
   ii. How many different routes could a ball take to Column 5? What is the probability of a ball falling into Column 5?

16 In Problems 5 and 6 of Investigation 2, you made histograms of binomial distributions. For the softball context, \( n = 5 \) and \( p = 0.4 \). For the jury pool situation, \( n = 12 \) and \( p = 0.3 \). The graphs below for these two situations were produced using the TCMS-Tools “Binomial Distributions” custom app under the Statistics menu. Use this computer software or similar software to explore the following questions.
ON YOUR OWN

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

b. As \( p \) increases from 0.01 to 0.99 but \( n \) remains the same (for example \( n = 5 \)), what happens to the shape, center, and spread of the binomial distribution for the number of successes?

17 According to the United States Census Bureau, the 2010 population of Washington, D.C., was approximately 51% black. (Source: www.census.gov/prod/cen2010/briefs/c2010br-06.pdf)

a. Suppose a jury of 12 people was picked at random from the population of Washington, D.C. What is the probability that all of the jurors were black?

b. Suppose that 10 juries were picked at random from the population of Washington, D.C. What is the probability that at least one jury consisted of all black members?

18 According to the United States Census Bureau, the 2010 population of the state of Washington was approximately 11% Hispanic. (Source: www.census.gov/prod/cen2010/briefs/c2010br-04.pdf)

a. Suppose a jury of 12 people was picked at random from the population of Washington. What is the probability that none of the jurors were Hispanic?

b. Suppose that 10 juries were picked at random from the population of Washington. What is the probability that at least one jury consisted of all non-Hispanic members?

REFLECTIONS

19 Suppose that you flip a fair coin ten times.

a. Without computing, are you more likely to get 4 heads or 3 heads, or are the probabilities the same? Explain your reasoning.

b. Are you more likely to get 1 head or 9 heads, or are the probabilities the same? Explain your reasoning.

c. List a sequence of five heads and five tails that you could get. What is the probability of this particular sequence?

d. List a sequence of nine heads and one tail that you could get. What is the probability of this particular sequence?

e. If you flip a fair coin ten times, are you more likely to get five heads and five tails or nine heads and one tail? How can you reconcile this with your answers from Parts c and d?
ON YOUR OWN

20 Look back at Applications Task 6. If you used a technology tool in completing that task, what tool did you use and how did you make the decision on tool use? If you did not use technology, reflecting back on the task, what might be an appropriate technology tool to use? Why?

21 About 24% of residents of the United States are under the age of 18. (Source: www.census.gov/prod/2010br-03.pdf)
   a. If you select 10,000 United States residents at random, what is the expected number of them who are under the age of 18?
   b. Use technology to find the probability of getting exactly 2,400 residents who are under the age of 18.
   c. Suppose that you take a random sample of 10,000 residents from your state and exactly 2,400 of them are under age 18. Why doesn’t the result of Part b mean that you have a statistically significant result? What is the correct P-value?

22 In New York City, 34% of people over age 24 are college graduates. (Source: quickfacts.census.gov/qfd/states/36000.html)
   a. Would it be statistically significant if a jury has 0 college graduates? Only one college graduate?
   b. Suppose a lawyer in New York City keeps track of the number of college graduates on the next 1,000 juries. She finds that 25 of the 1,000 juries have so few college graduates as to qualify as statistically significant. She plans to recommend that the jury selection process be investigated. What would you say to her?

EXTENSIONS

23 This task involves situations where sampling is done without replacement and the sample size is more than 10% of the size of the population.
   The graduating class at Smallville High School has 10 students. Seven of them already have their graduation robes. You plan to randomly select two different students for two “candid” yearbook photos and would like both students to be photographed in their graduation robes.
   a. What is the probability that the first student you select has his or her robe?
   b. Suppose that the first student selected has his or her robe. What is the probability that the second student also has his or her robe?
   c. Compute the probability that both students have their robes.
   d. Now suppose that you will select four students at random. Compute the probability that they all have their robes.
24 Chaser is a remarkable border collie who knows the names of more than a thousand objects. To train Chaser, a sample of 20 objects was randomly selected from those Chaser had learned. They were placed randomly around the floor in a different room from where Chaser and the trainer were waiting. The trainer was given a randomized list of the 20 objects and asked Chaser to retrieve the first object listed. When Chaser returned, the trainer continued down the list, without replacing each object, until Chaser had retrieved all 20 items. (Source: John W. Pilley and Alliston K. Reid, “Border collie comprehends object names as verbal referents,” Behavioural Processes, Vol. 86 (2011) pages 184–195.)

a. If Chaser is clueless about the first item and selects one at random to bring back to the trainer, what is the probability she selects the correct item?

b. If Chaser is clueless about each item, what is the probability that she gets all 20 objects correct, just by chance? (She almost always did this.)

c. How would your answer to Part b change if there were still 20 trials but each object was returned to the room before Chaser was asked to retrieve another?

25 Refer to Extensions Task 23. In a large city, 1,465,241 of the 2,000,000 households recycle soft drink cans. If five households are selected at random, what is the probability that all five households recycle:

a. if the households are selected without replacement (all five households must be different)?

b. if the households are selected with replacement (you can select a household more than once)?

c. Does it make much difference in your answers whether the households are selected with replacement or without replacement?

26 Look back at Problem 9 Part d (page 352) where you considered how expected value can be used to help price insurance. Prove that the break-even point for the insurance company occurs at a price of $1.29 per policy regardless of the number of policies sold.
ON YOUR OWN

27 She thought he was cheating on her. So, Lena Sims Driskell, age 78, shot and killed her boyfriend, age 85. “During jury selection, defense attorney Deborah Poole complained that Mrs. Driskell could not receive a fair trial with a jury of her peers because the juror pool lacked enough older people from which to select.” Of the 58 people in the jury pool, all but five appeared to be younger than age 65. (Source: www.legalzoom.com/crime-criminals/murder/do-you-have-right and www.gainesville.com/apps/pbcs.dll/article?AID=/20060620/WIRE/60620020/1117/news)

a. Can you use the binomial probability formula to compute the probability that a jury of 12 people that is randomly selected from this pool would contain no one who appeared to be age 65 or older? Explain.

b. Use an appropriate method to find the probability that a jury of 12 people who are randomly selected from this jury pool would contain no one who appeared to be age 65 or older.

c. What can you conclude?

28 Acceptance sampling is one method that industry uses to control the quality of the parts it uses or other products for manufacturing. For example, a company that produces organic vegetable juices regularly receives shipments of vegetables from a supplier. To ensure the quality of the vegetables, an employee examines a sample of the vegetables in each shipment. The shipment is accepted if only 5% or fewer of the vegetables in the sample are considered low-quality. Assume that 10% of the vegetables from this supplier are low-quality. Suppose that the employee examines a random sample of 20 vegetables from each shipment.

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

b. Design and carry out a simulation of this situation for 200 random samples.

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

d. If the company wishes to reduce the probability that they accept a shipment with 10% low-quality vegetables, what should they do?

29 Suppose that a population of size \( N \) contains \( S \) successes and \( F \) failures. You take a random sample of size \( n \) from this population.

a. If the sampling is done with replacement, explain why the probability of getting exactly \( s \) successes and \( f \) failures is given by the following formula:

\[
C(n, s) \left( \frac{S}{N} \right)^s \left( \frac{F}{N} \right)^f
\]

b. If the sampling is done without replacement, explain why the probability of getting exactly \( s \) successes and \( f \) failures is given by the hypergeometric formula:

\[
\frac{C(S, s) \cdot C(F, f)}{C(N, n)}
\]

c. Suppose a group of 25 students includes 15 students who play a musical instrument. If the sampling is done with replacement, find the probability that if you select 8 students at random, exactly 6 play a musical instrument. What is the probability if the sampling is done without replacement?
ON YOUR OWN

d. A high school football team has 40 players, half of whom are seniors. The coach has eight extra tickets to the homecoming game. He will select eight different players at random to get a ticket. What is the probability that exactly half will be seniors?

e. Refer to Extensions Task 27. Use the appropriate formula to find the probability that a randomly selected jury would have no one on it who appeared to be age 65 or older.

REVIEW

30 In previous studies, you used statistics appropriate to the context and the shape of a distribution to describe the center (median, mean) and the spread (interquartile range, standard deviation) of the distribution. You represented the data with dot plots, histograms, and boxplots. You may have recognized that the mean of a distribution displayed as a dot plot or histogram is the balance point of the distribution. Below is a dot plot of the dissolution times from a chemistry experiment. Students measured the time in seconds for a solute to dissolve. Study the plot to answer the questions below.

![Dot plot of dissolution times](image)

a. How many dissolution times were collected by the science students?
b. Make a quick, rough estimate of the mean of the distribution from the dot plot. Then calculate the mean of the distribution.
c. What percentage of the dissolution times were less than 9 seconds? More than 16 seconds?
d. What percentage of the dissolution times are between 9 and 16 seconds, including 9 and 16?
e. Make a rough estimate from the dot plot of the percentage of dissolution times that are within 2 seconds of 12 seconds. Then calculate that percentage.

31 Rewrite each expression in equivalent form as a single algebraic fraction. Then simplify the result as much as possible.

a. \(-a + \frac{1}{a}\)  
b. \(\frac{a}{b} - 1\)

c. \(\frac{b}{a} + \frac{a}{b}\)  
d. \(\frac{z(z - 1)}{z}\)

e. \(\frac{x}{y} - \frac{1}{xy}\)  
f. \(\frac{1}{x - 1} - \frac{1}{x + 1}\)
32 Suppose you are asked to figure new prices for items in a music and video store. Show two ways to calculate each of the following price changes—one that involves two operations (either a multiplication and an addition or a multiplication and a subtraction) and another that involves only one operation (multiplication).

a. Reduce the price of a $50 video trilogy set by 20%.

b. Increase the price of a $10 CD by 30%.

c. Reduce the price of a $15 CD by $3.33.

33 The population of a fruit fly colony is given by the function \( p(t) = 16(2^t) \), with \( t \) representing time in days since the start of the experiment.

a. What is the population of the fruit fly colony after three days?

b. How long will it take for the population to reach 1,000 fruit flies?

34 Rewrite each expression in standard polynomial form.

a. \((3x - 4)^2 - 6\)

b. \(-3x(2x^2 - 1) - 8x + 7\)

c. \(8x - 4x^2 + 2x(3x - 4)\)

d. \((3 - 4x^2)(3 + 4x)\)

35 The Clifton Public Library surveyed 1,232 randomly selected adults living in the city about their reading habits. The results of part of their survey are summarized in the table below.

<table>
<thead>
<tr>
<th>Read a Print Book in the Last Year</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>466</td>
<td>495</td>
<td>961</td>
</tr>
<tr>
<td>No</td>
<td>64</td>
<td>207</td>
<td>271</td>
</tr>
<tr>
<td>Total</td>
<td>530</td>
<td>702</td>
<td>1,232</td>
</tr>
</tbody>
</table>

Suppose that you interviewed a randomly selected person from the sample.

a. What is the probability that the person had read both a print book and an ebook in the last year?

b. What is the probability that the person had read an ebook during the last year?

c. What is the probability that the person had read a print book during the last year?
ON YOUR OWN

d. Based on your answers to Parts a–c, are reading ebooks and print books independent for the people in this sample?

e. Perform a chi-squared test of independence to determine if there is statistically significant evidence that these two variables are not independent in the population of adults living in Clifton.

36 Write a function rule that matches the transformation of the graph of the function \( y = 2^x \).

a. Translated three units up

b. Translated four units to the right

c. Point \( A(0, 1) \) is mapped to point \( A'(0, 5) \) and point \( B(1, 2) \) is mapped to point \( B'(1, 10) \).

d. Point \( S(0, 1) \) is mapped to point \( S'(0, -1) \) and point \( T(1, 2) \) is mapped to point \( T'(1, -2) \).

37 Solve each equation or inequality. Display solutions for inequalities in interval notation.

a. \( 3x + 4 = 20 \)

b. \( 70 - \frac{3}{4}x < 10 \)

c. \( \frac{x + 1}{3} = \frac{2x - 1}{2} \)

d. \( 45 \leq -5x + 10 \)

38 Solve each equation for \( x \).

a. \( x^2 - x - 6 = 0 \)

b. \( 3x^2 - 1 = 47 \)

c. \( 16x^3 - x = 0 \)

d. \( 5x^2 - 13x = 6 \)
Sample Surveys

Surveys of samples of people are used often by government agencies, the media, and consumer-oriented businesses to better understand characteristics of the American people, their behavior, and their preferences. Political polls assess public opinion about issues and candidates. A large fraction of these surveys are worthless, but surveys from the U.S. government and reputable companies do a very good job. You may have wondered how they manage to measure public opinion accurately without getting everyone’s opinion.

The Gallup company conducts surveys to monitor various characteristics of Americans. For example, it periodically asks a sample of people a series of questions about their lives and then classifies each person as thriving, struggling, or suffering. This survey is conducted by phone with about 28,295 randomly selected American adults. In October 2012, 51% of the Americans surveyed were classified as “thriving,” 45% as “struggling,” and 4% as “suffering.”

(Source: www.gallup.com/poll/158543/americans-life-outlook-better-2008-not-best.aspx)
THINK ABOUT THIS SITUATION

Think about the conduct and findings of the Gallup poll described on the previous page.

a. Why do you think this survey is given to a sample of Americans instead of getting information from all Americans?

b. How can these percentages be trusted when only 28,295 Americans were surveyed?

c. What other surveys have you read or heard about?

d. What background information about a survey do you think is important to know in order to decide whether you can trust it?

In this lesson, you will learn how surveys and polls can be conducted so that the results are trustworthy.

INVESTIGATION 1

Trustworthy Surveys

People take surveys for many different reasons. Government surveys are taken to investigate the jobless rate and the cost of living. Opinion polls are surveys used to gauge public opinion on political issues. Some of these surveys are completely trustworthy. That is, you can be pretty sure that the percentages reported are fairly close to those that would have resulted had the entire population been surveyed. On the other hand, the results from some surveys are not worth anything. You will learn about the first kind of survey in this investigation and about worthless surveys in the following investigation.

As you work on the problems in this investigation, look for answers to this question:

What are the characteristics of a trustworthy survey?

1. Individual opinions can be collected by using a census or by using a sample survey. A census collects information from every individual in a population, the entire set of people or things you would like to describe. In a sample survey, questions are asked only of a subset of a population. If the respondents are selected at random, a survey can accurately assess the characteristics of the entire population without contacting every individual. Read the following information from a trustworthy sample survey that estimated the proportion of people age 14–24 in the United States who have been cyberbullied.
In this lesson, you will learn how surveys and polls can be conducted so that the results are trustworthy.

Trustworthy Surveys

People take surveys for many different reasons. Government surveys are taken to investigate the jobless rate and the cost of living. Opinion polls are surveys used to gauge public opinion on political issues. Some of these surveys are completely trustworthy. That is, you can be pretty sure that the percentages reported are fairly close to those that would have resulted had the entire population been surveyed. On the other hand, the results from some surveys are not worth anything. You will learn about the first kind of survey in this investigation and about worthless surveys in the following investigation.

As you work on the problems in this investigation, look for answers to this question:

What are the characteristics of a trustworthy survey?

1. Individual opinions can be collected by using a census or by using a sample survey. A census collects information from every individual in a population, the entire set of people or things you would like to describe. In a sample survey, questions are asked only of a subset of a population. If the respondents are selected at random, a survey can accurately assess the characteristics of the entire population without contacting every individual. Read the following information from a trustworthy sample survey that estimated the proportion of people age 14–24 in the United States who have been cyberbullied.

**INVESTIGATION**

**Cyberbullying Widespread, MTV/AP Survey Reveals**

“More than half of the respondents to an MTV/Associated Press study revealed they have been the targets of mean behavior or fake gossip on social-networking sites or text messages. The pervasiveness of digital abuse and cyberbullying in the study confirms a disturbing trend in which young people are using the Internet and wireless devices to harass each other, but also reveals that more are stepping up and saying something when they see abuse online. According to the study, 76 percent of 14–24 year olds say digital abuse is a serious problem for people their age, with 56 percent reporting that they have experienced abuse through social and digital media.”

The study was conducted by Knowledge Networks using 1355 participants, age 14–24, chosen scientifically by a random selection of telephone numbers and residential addresses. Sampling margin of error for a 50% statistic with 95% confidence is ±3.8 for all interviews. The survey has a response rate of about 65%.


a. Was this a census or a sample survey?

b. Surveys often are used to estimate a numerical characteristic of the population, called a parameter. For example, the percentage of “successes” in a sample can be used as an estimate of the percentage of successes in the population (the parameter).

i. Describe the population that was being studied in the cyberbullying research.

ii. What parameter is being estimated by 56%?

iii. Find the statement that gives the margin of error, which tells about how far off that estimate might reasonably be. What is the margin of error?

c. Was the sample selected at random? How large was the sample?

d. What additional information would you want to know in order to decide whether the estimate of the percentage of all people age 14–24 who would say they have been cyberbullied is trustworthy?
Describe the population in each case. Then, decide whether you would conduct a census or sample survey to answer each of the questions below.

a. How do students in your class feel about a particular new movie?

b. Is a manufacturer producing a high percentage of light bulbs that are defective?

c. How many students bought a hot lunch at your school today?

d. Who are people planning to vote for in the next election for governor of your state?

The results of a survey may be interesting, but before you trust the results, it is important to evaluate how the survey was constructed and carried out. Here are questions you should ask:

1. What is the issue of interest or the parameter being estimated?
2. What is the population?
3. How was the sample selected?
4. How large is the sample?
5. What was the response rate (percentage of people contacted to be in the sample who gave a response)?
6. How were the responses obtained from the sample (personal interview, phone interview, Internet, etc.)?
7. What were the exact questions asked?
8. Who sponsored the survey? Did a reputable company conduct the survey?

Refer to the study in Problem 1. Which of the questions above cannot be answered based on the information given?

A trustworthy survey always selects the participants at random. There are two main ways to do this:

- **(Simple) random sampling** is done by a process equivalent to writing the people’s names on identical slips of paper, mixing them up, and then selecting slips of paper at random. In a random sample of size \( n \), each possible sample of \( n \) people has an equal chance of being the sample.

- **In stratified random sampling**, the population first is divided into non-overlapping strata, or groups of similar people, such as men and women or Republicans, Democrats, and other. Then a random sample is taken from each stratum. The sizes of the random samples usually are proportional to the sizes of the strata. For example, suppose you want to take a stratified random sample of 1,500 households from a population in which 42% contain children under the age of 18 and 58% do not. First select a random sample of 0.42(1,500) = 630 households from those with children. Then select a random sample of 0.58(1,500) = 870 households from those without children.
Consider the following methods of getting a sample of 50 teen drivers from a town.

**Method I**
Go to the high school, ask for a list of students, and select 50 at random. Replace each student who is not a driver with a different randomly selected student until you have 50 drivers.

**Method II**
Contact the Department of Motor Vehicles, get a list of all drivers in the town under age 20. Select 50 at random.

**Method III**
Contact the Department of Motor Vehicles, get a list of all drivers in the town under age 20. Choose a person at random, and then select that person and the next 49 whose names appear on the list.

a. Which method produces a random sample?

b. Why do the other two methods not produce a random sample?

In a typical pre-election poll, probable voters are stratified by gender, by political party preference, and by race.

a. Why might polling organizations want to use a stratified random sample rather than a simple random sample?

b. In the United States during the week of September 24th, 2012, about 28% of adults considered themselves Republicans, 32% considered themselves Democrats, 38% considered themselves Independents, and the rest were undecided. [Source: www.gallup.com/poll/15370/Party-Affiliation.aspx] Suppose you were taking a political poll that week of 1,200 adults. You wanted to stratify on party preference so that the sizes of the samples were proportional to the sizes of the strata. How many adults should you sample from each group?

Suppose you are interested in student opinion about a proposed new school mascot.

a. Describe how you would select a random sample of students from your school.

b. How might you use stratification to be sure that every important group of students is represented in your survey?

**SUMMARIZE THE MATHEMATICS**

In this investigation, you found that when you read about a survey used to measure public opinion, it is important to evaluate how much you can trust the results.

a. Describe the difference between taking a sample survey of the students in your school and taking a census of the students.

b. What is meant by the “population” when you take a survey? What is a population parameter?

c. What is the best method of getting a sample? Describe the two main methods of doing this.

d. What questions should you ask about the design and conduct of a survey in order to completely understand and accept the results?

*Be prepared to share your ideas and reasoning with the class.*
Pei has surveyed students about whether they favor the block scheduling that was introduced at their school last fall. To ensure privacy about this contentious issue, he mailed surveys, which could be returned anonymously, to the homes of students in his sample. He could afford to mail the survey to 150 students. Pei believed that the response would vary depending on the student’s class year, so he stratified on class year, with the sizes of the random samples proportional to the sizes of the strata. His high school has 450 freshmen, 379 sophomores, 412 juniors, and 352 seniors. Of the 150 surveys he sent out, 135 were returned. Of these, two-thirds favored the new block scheduling.

a. To how many freshmen did Pei send a survey? Sophomores? Juniors? Seniors?

b. Refer to the questions in Problem 3 on page 378. Answer the questions that can be answered based on the information given.

c. How trustworthy do you find Pei’s survey?

**INVESTIGATION 2**

**How Surveys Can Go Wrong**

When interpreting survey results, it is important to consider how the survey was constructed and carried out. In this investigation, you will examine how bias may occur in a survey.

As you work on the problems in this investigation, look for answers to this question:

*What are the characteristics of an untrustworthy survey?*

1. As most people realize, the estimate from a sample probably will not be equal to the population parameter. Two possible sources of error are chance and bias.

   - **Chance (or sampling) error** results from the fact that a survey based on a random sample does not ask everyone in the population. Thus, the estimate from the sample may not be exactly equal to the population parameter. In Lesson 3, you will learn that a larger sample size tends to reduce sampling error.

   - **Bias**, on the other hand, tends to push the estimate to one side of the population parameter. Specifically, in repeated sampling, the estimate from the sample is too big or too small, on average.
In the diagrams below, the center of the target represents the population parameter. The ×s represent estimates from the sample, attempts to hit the center of the target.

a. Which target shows errors due to chance alone? Which shows error due to both chance and bias?
b. Draw a target that shows error due to bias, with very little chance error.
c. Draw a target that shows almost no error due to bias or to chance.
d. Share your targets and explanations with others. Resolve any differences.

2 Shown below is a reduced copy of the set of circles shown on the following page.

Your task in this problem is to estimate the average area of all 115 circles shown on the following page. It certainly would be a lot of work to compute the area of each circle, so this is a situation where sampling might be better. You will be comparing two different methods to get your sample.

a. First use a judgment sample, using your best judgment to select five circles for the sample. That is, select five circles you think are fairly typical. Record the radius and area of each of your five circles. Finally, compute the average of the areas of the five circles.
b. Next, use technology to generate five random numbers between 1 and 115 inclusive and locate the corresponding circles. Find the average area of the five randomly selected circles.
c. Now, pool your data with other students in your class, making two lists and two dot plots for the averages generated in Parts a and b. Compare means of the two lists. Are the values about equally spread out in the two plots?
d. Check with your teacher on the actual average area of all 115 of the circles. Compare that area to the means computed in Part c. Was anything surprising in this comparison? Can you make any conjectures about how samples should be selected?
e. A method of selecting a sample is **biased** if repeated samples would yield estimates that are systematically too large (or too small), on average. Types of **sample selection bias** (bias that results from the method of selecting the sample) include
   - **size bias**: When using their own judgment, people tend to pick the larger items to be in the sample.
   - **voluntary response bias**: Allowing people to decide on their own whether to be in the sample generally results in a sample consisting of people with strong opinions about the issue.
   - **convenience sample**: Using a readily available group as the sample typically results in a sample with views that are not representative of the population.
   - **wrong population**: The sample was not selected from the population of interest.

Do you think either of the methods of selecting a sample of five circles is biased? If so, which type of bias did it illustrate?

f. Random sampling is used in surveys because it eliminates all but one of the types of bias above. Which one does it not eliminate?

3 The most famous polling mistake in U.S. history changed the way that polls are conducted. This poll had a huge sample, but went wrong anyway. In 1936, the **Literary Digest** conducted a poll of millions of voters and declared that Republican Alfred Landon would win the presidential election. Instead, Franklin Roosevelt won by a landslide. The magazine had sent out ballots printed on more than 10,000,000 postage-paid postcards and over 2,300,000 were returned. The names of people who received the postcards came mostly from **Literary Digest** subscribers, automobile registration lists, and telephone directories. (Source: Squire, Peverill. "Why the 1936 Literary Digest Poll Failed." The Public Opinion Quarterly, Vol. 52, No. 1 (Spring 1988), pages 125–133.)

a. What was the response rate in the **Literary Digest** survey?

b. What types of sample selection bias may have occurred as a result of the sampling procedure used?

c. George Gallup was able to predict the result of the election correctly, with a sample of only 3,000 voters. How do you think he was able to do this?
Surveys also can go wrong because the pollster gets no response or an incorrect response from a perfectly good random sample. Types of response bias (bias that results from systematically incorrect responses) include:

- **nonresponse bias**: People who have been selected for the sample cannot be contacted or refuse to answer.
- **questionnaire bias**: The question is worded so as to lead people to one particular answer or worded so that people in the sample misunderstand the question.
- **incorrect response**: People give the wrong answer. For example, maybe they do not remember accurately or they do not want the interviewer to know what they really think.
- **bad timing**: The time the survey was conducted influenced many people’s response.
- **measurement error**: A mistake is made in collecting or analyzing the results. For example, a telephone interviewer incorrectly records a person’s response or a person accidentally fills in the incorrect response on a written questionnaire.

4 What kind of bias may have occurred in the surveys described below? Do you think that as a result of the bias the estimate from the sample will be larger or smaller than the parameter?

a. Every year, the National Center on Addiction and Substance Abuse at Columbia University surveys a random sample of teens and parents by phone. One question they ask teens is, “Is your school a drug-free school or is it not drug free, meaning some students keep drugs, use drugs or sell drugs on school grounds?” In the 2012 survey, 53% of the teens surveyed answered “Drug free.” To check on possible response bias, teens were asked one final question, “As you were speaking with me, was there someone there with you who could overhear your answers?” Twenty-two percent of the teens answered “yes” to the last question. (Source: www.casacolumbia.org/templates/publications_reports.aspx)

b. Cali says, “The results of my survey showed that 21% of the boys and 30% of the girls support me for president of the senior class. So, I should win the election with 51% of the vote.”

c. A university sent out a survey to all alumni to find out how alumni felt about the usefulness of their education. Twenty-seven percent of the alumni returned the survey. The responses were overwhelmingly positive.
d. Harris Interactive conducted an online survey of 2,331 U.S. adults, asking, “If you had to choose, which one of these sports would you say is your favorite?” The sport chosen most often was professional football (31%). The survey was conducted in the middle of December. (Source: www.harrisinteractive.com/vault/HI-Harris-Poll-Favorite-Sport-Football-2011-01-20.pdf)

e. Student.com is an online community for teens and college students. It has a “Poll of the Day” in which members can respond to a question. Recently, the question was “How many piercings do you have?” Of the 363 students who responded, 37.2% said “None.” (Source: www.student.com/potdmore.php?id=406)

5 The Panama Canal, opened in 1914, was built by the United States.

![Image of the Panama Canal]

In 1978, by one vote, the U.S. Senate ratified a treaty that turned over the operation of the canal to the country of Panama. The Senate was influenced by polls taken to gauge public opinion on this issue. The wording of three questions used by various polls is given below. (Source: Ted J. Smith III; J. Michael Hogan, Public Opinion and the Panama Canal Treaties of 1977, The Public Opinion Quarterly, Vol. 51, No. 1. [Spring, 1987], pages 5–30.)

**Question 1:** In September, the Presidents of the United States and Panama signed two treaties which would gradually turn the Panama Canal over to the Panamanians but would provide for the continued use and defense of the Canal by the United States. Before these treaties can take effect, the United States Senate must act on it. Do you think the Senate should vote for the new treaties or against them?

**Question 2:** Do you favor the United States continuing its ownership and control of the Panama Canal or do you favor turning ownership and control of the Panama Canal over to the Republic of Panama?

**Question 3:** Do you think the time has come for us to modify our Panama Canal treaty or that we should insist on keeping the treaty as originally signed?

a. Which question do you think has the best wording? Explain.

b. For these questions, the percentage of those surveyed who gave an answer that supported the proposed treaty were—not necessarily in order—13%, 30%, and 45%. For which question do you think 45% supported the treaty? Explain your thinking.

c. Which type of response bias does this illustrate?
SUMMARIZE THE MATHEMATICS

In this investigation, you learned that sample selection is an important part of conducting a trustworthy survey. Equally important is getting a good response from each person selected for the sample.

a. Explain why random sampling is a good method for sample selection.

b. What are some types of sample selection bias?

c. What are some types of response bias?

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

CHECK YOUR UNDERSTANDING

For each study below, complete Parts a, b, and c to consider whether or not the survey method was trustworthy.

Study 1  A contemporary hit (top 40) radio station wants to estimate the percentage of voters who are in favor of raising the age that people in the state can get a driver’s license. It asks its listeners to call in and give their opinion. The sample consists of those who call in on a toll-free number and state their opinion.

Study 2  A psychologist selects a random sample of 60 students from your school for a study about emotion. They are taken to the auditorium where the psychologist asks the students to raise their hand if they have ever cried during a movie.

Study 3  The town council wants to find out what proportion of citizens are bothered by barking dogs. It mails a survey to each citizen asking them to fill it out and return it by mail.

Study 4  An employer personally asks all employees if they think his company’s policy about vacation time is fair.

Study 5  The homecoming committee wants to know the percentage of students who plan to buy a ticket to the homecoming dance. They ask all of the students in each of their classes whether they plan to buy a ticket.

a. Identify any possible bias in the method of selecting the sample.

b. Identify any possible response bias.

c. Do you think the estimate is likely to be too high or too low? Explain your reasoning.
APPLICATIONS

1 Read the following summary of a sample survey.

**U.S. Religious Knowledge Survey**

Researchers from the independent Pew Forum on Religion & Public Life phoned 3,412 Americans and asked them questions about various religions. Atheists and agnostics, Jews, and Mormons scored the highest. This result cannot be attributed to age or educational level. Here is one question that the interviewer read to the respondents. The choices were to be randomized for each respondent.

Please tell me which of the following is NOT one of the Ten Commandments:

• Do not commit adultery.
• Do unto others as you would have them do unto you.
• Do not steal.
• Keep the Sabbath holy.
• All are in the Ten Commandments.

Fifty-five percent answered this question correctly.

Results for this survey are based on telephone interviews conducted under the direction of Social Science Research Solutions (SSRS) among a national sample of 3,412 adults living in the continental United States, 18 years of age or older, from May 19-June 6, 2010 (2,393 respondents were interviewed on a landline telephone, and 1,019 were interviewed on a cell phone, including 444 who had no landline telephone). The survey of the full national population used “random digit dial” (RDD) methodology. Interviews were conducted in English and Spanish. The overall response rate for this study is 17.2%.


a. For the question given in the summary, what is the population being studied? Describe the parameter being estimated. What is the estimate of this parameter?

b. What was the method of selecting the sample? How large was the sample? What is the response rate?

c. How were the responses obtained? Is the exact question given?

d. Does there appear to be any reason to find this poll untrustworthy? Explain.

2 Find an example of a survey on TV, in the newspaper, or on the Internet.

a. Write a brief summary of the survey.

b. Referring to the questions in Problem 3 of Investigation 1 (page 378), answer the questions that can be answered based on the information in the media-reported survey.
Wareham High School, in Massachusetts, has about 900 students in grades 9–12. Read the following information about a survey to assess student opinion about mandatory school uniforms.

Uniforms May Be Gaining Momentum for Wareham Schools
by Laura Fedak Pedulli
November 9, 2012, 12:00 AM

WAREHAM —The prospect of uniforms at the town’s public schools may be gaining momentum.

According to results of a student-wide survey at Wareham High School obtained by The Standard-Times, while many objected to uniforms, 98 of the 184 surveyed expressed a willingness to explore the possibility and 22 students said they would consider working on a uniforms committee or help design the potential new look.

The survey was administered in June by Assistant Principal Debbie Freitas but results weren’t tallied until Monday. ...

Both Freitas and Principal Scott Paladino also said uniforms would help with enforcing the current dress code as expected attire would be firmly defined.

Freitas stressed that students’ willingness to embrace a uniform policy would be key to whether the idea takes off. ...

Source: www.southcoasttoday.com/apps/pbcs.dll/article?AID=/20121109/NEWS/211090339/-1/NEWS10

a. Referring to the questions in Problem 3 of Investigation 1 (page 378), answer the questions that can be answered based on the information in the abstract above.

b. Does there appear to be any reason to find this survey untrustworthy? Explain.

c. Does the Assistant Principal have statistically significant evidence that a majority of students would have expressed a willingness to explore the possibility of uniforms, if she had asked all of them? Select the best choice in each pair of options below.

Assuming that the proportion of all students who would have expressed this willingness is exactly/more than 50%, we compute $1 - \text{binomcdf}(184,0.50,97) \approx 0.2087$. Because this $P$-value is more/less than 5%, the Assistant Principal does/does not have statistically significant evidence that the proportion of all students who are willing is exactly/greater than 50%. In other words, if the true proportion is 50%, it would/would not be very likely to get 50/53% or even more who are willing in a sample of size 98/184.
4 Yen was quite upset about the way the Senior Class Council was planning graduation. She sent a survey to all members of the senior class. A reporter for the high school newspaper asked the AP Statistics class to evaluate Yen's survey and then wrote the following article based on their report.

Survey About Senior Class Council Found To Be Full of Flaws
by Julie Adalian

Last month’s survey of the senior class blasted the work of the Senior Class Council on planning graduation. About three-quarters of the seniors who returned the survey agreed that the Senior Class Council was incompetent.

Following the uproar, Mr. McCune’s AP Statistics class analyzed the survey as part of a unit in the design and analysis of surveys. Their review concluded that the survey was “biased,” “deeply flawed,” and “no valid conclusion could be drawn from it.” They said that the survey was a clear attempt to force an election to replace current members of the Senior Class Council.

Yen Chang, a senior who was a candidate for the Senior Class Council, but was not elected, personally handed a copy of the questionnaire to each member of the senior class, asking each person to return it to her. She included a “fact sheet” with the survey that spelled out why she thinks the Senior Class Council is incompetent. She detailed why she thought the planning process for graduation was “amateur, incomplete, wasting our money, and behind schedule.”

Mr. McCune’s class noted that only 32% of the seniors returned the survey and many of those who did were students in Yen’s classes.

Another problem with her survey was lack of confidentiality. Mr. McCune’s class suggested that next time such a survey be returned to Mr. Gonzales in the main office, so that students wouldn’t be concerned about another student knowing how they responded. Finally, it appears that Yen made some arithmetic mistakes in compiling her results.

Yen defended her survey saying, “Anyone who wanted to could return the survey to me. I took them all, not just those of my friends. I don’t think that the fact sheet contained information that most people hadn’t heard before, but I thought it important to remind them so they could make an informed decision.”

a. Referring to the questions in Problem 3 of Investigation 1 (page 378), answer the questions that can be answered based on the information in the article above.

b. List each source of bias that is mentioned. Classify each source of bias as sample selection bias or response bias.

c. Do you agree that this survey is untrustworthy? Do you think that the result overestimated or underestimated the proportion of seniors who think the Senior Class Council is incompetent? Explain your answers.
5 In the 1948 presidential election, Harry Truman was running against Thomas Dewey. All of the major polling organizations of the day had learned not to make the same mistakes as made by the Literary Digest poll of 1936 (see Problem 3 on page 378). Yet, they still were mistaken about who would win the election.

A good explanation of what went wrong can be found at the Web site of the Truman Library: www.trumanlibrary.org/whistlestop/study_collections/1948campaign/large/docs/victory_final/what_happened.htm. Go to the Web site and read the analysis, starting with the paragraph that begins, “Everybody’s mistakes, in the end, could be attributed to the polls.”

a. List each source of bias that is mentioned.

b. Classify each source of bias as sample selection bias or response bias.

6 In a classic 1940 experiment about the wording of questions, respondents were asked one of the following two questions.

**Question 1:** Do you think the United States should allow public speeches against democracy?

**Question 2:** Do you think the United States should forbid public speeches against democracy?

(Source: Rugg, D. Experiments in wording questions: II. Public Opinion Quarterly 5 (1941) pages 91–92.)

a. Is there any logical difference between the two questions?

b. Which question do you think resulted in a larger proportion of people who did not want to allow speeches against democracy? Explain your reasoning.

7 The three questions below about the scientific evidence for evolution have been asked to large random samples of Americans. (Source: pollingreport.com/science.htm)

**Question 1:** Do you think the scientific theory of evolution is well-supported by evidence and widely accepted within the scientific community?

**Question 2:** Just your opinion: Do you think that Charles Darwin’s theory of evolution is a scientific theory that has been well-supported by evidence, or just one of many theories and one that has not been well-supported by evidence, or do you not know enough about it to say?
**Question 3:** Last year the National Academy of Sciences recommended that evolution be taught to all public school students as the most convincing theory for how human beings developed. Do you agree or disagree that evolution should be taught in all public schools?

a. Rank these questions in order from the one you think resulted in the most support for the scientific evidence for evolution to the one you think had the least support.

b. Which question do you think has the best wording? Explain.

**CONNECTIONS**

Suppose that, in your school, 30% of the students are freshman, 28% are sophomores, 25% are juniors, and 17% are seniors.

You plan to survey a total of 300 students to estimate what percentage of students favor an increase in the size of the student council. If you want a stratified random sample with the sizes of the samples proportional to the sizes of the strata, how many students should you select from each class year?
9 In the United States, there are 2,816,000 high school dropouts, who have not completed high school and are not enrolled in school, between the ages of 16 and 24. Of these, 1,290,000 are employed, 527,000 are unemployed, and 999,000 are not in the labor force. (Source: www.census.gov/compendia/statab/2012/tables/12s0274.pdf)

High School Dropouts

a. Suppose a pollster wants to conduct a survey with a stratified random sample of a total of 1,500 dropouts. How many should she select from each group if she wants the sizes of the samples to be proportional to the sizes of the strata?

b. In her sample, the pollster finds that 65% of those employed, 80% of those unemployed, and 32% of those not in the labor force have plans to go back to school. What is her best estimate of the total percentage of dropouts who plan to go back to school?

10 The population of the world is about 33% Christian, the largest percentage of any religion. (Source: CIA Factbook; https://www.cia.gov/library/publications/the-world-factbook/) Suppose you take a random sample of 1,000 people from Ethiopia. Six hundred of the people in your sample are Christian.

a. Assuming that Ethiopia has the same percentage of Christians as the rest of the world, compute the P-value for this situation. Is this a statistically significant result?

b. What can you conclude?

11 In Problem 2 of Investigation 2 (pages 381–383), you considered the average area of a collection of circles. Think about some characteristics of “average circles.”

a. If a circle has a radius equal to the average radius of all circles in a collection, does it have a circumference equal to the average circumference? If so, explain why. If not, give an example of a small collection of circles such that a circle of average radius does not have average circumference.

b. If a circle has a radius equal to the average radius of all circles in a collection, does it have an area equal to the average area? Justify your answer.

c. What explains the difference in your answers to Parts a and b?
Below is a group of rectangles.

a. Show a copy of rectangles I–VII to 30 people and ask them: “Which rectangle do you prefer?” Write a brief summary of your findings.

b. A **golden rectangle** is a rectangle where the ratio of the length to the width is \(1 + \frac{\sqrt{5}}{2}\) or approximately 1.618.

   i. Which rectangle is closest to being a golden rectangle?

   ii. It has been claimed that people find the golden rectangle the most pleasing. (See the article at www.maa.org/external_archive/devlin/devlin_05_07.html for a discussion.) What proportion of your sample picked the golden rectangle?

   iii. If, overall, people have no preference among the rectangles, what proportion would pick the golden rectangle?
iv. Use the binomial probability function of your calculator to find the probability of getting as many people as you did in your sample, or even more, who would pick the golden rectangle just by chance.

v. Is your result statistically significant? What can you conclude?

The following question was asked of 506 adults nationwide as part of a special end-of-the-millennium poll.

Who do you feel is the greatest figure in the last thousand years, from anywhere in the world, specifically in the field of politics or government?

<table>
<thead>
<tr>
<th>Top Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>John F. Kennedy</td>
<td>13</td>
</tr>
<tr>
<td>Abraham Lincoln</td>
<td>12</td>
</tr>
<tr>
<td>Franklin D. Roosevelt</td>
<td>7</td>
</tr>
<tr>
<td>George Washington</td>
<td>6</td>
</tr>
<tr>
<td>Thomas Jefferson</td>
<td>5</td>
</tr>
<tr>
<td>Bill Clinton</td>
<td>4</td>
</tr>
<tr>
<td>Ronald Reagan</td>
<td>4</td>
</tr>
</tbody>
</table>


a. What is striking about these responses?

b. How could the question have been worded to give better responses?

The U.S. Bureau of Labor Statistics (BLS) tracks unemployment in the United States by conducting a monthly survey of 60,000 households. To get a better estimate of the unemployment rate, a question on the survey was reworded as follows.

Former question:

What were you doing most of last week?
- Working or something else?
- Keeping house or something else?
- Going to school or something else?

New question:

Last week, did you do any work for either pay or profit?

a. How might the former question lead to response bias?

b. What effect do you think revising the question will have on the estimate of the unemployment rate? Explain your reasoning.
Examine each of the following survey questions. For each survey question, explain how this question might lead to response bias. Then rewrite the question to eliminate the possible bias.

**Question 1:** This question was asked on an opinion poll about taxes: “Do you agree that the current high tax structure is excessive?”

**Question 2:** These instructions were given for an opinion poll about a new movie: “Rate the movie 1 to 10, where 1 is best.”

**Question 3:** To determine the number of people looking for jobs, an interviewer asked survey respondents, “Are you unemployed?”

**Question 4:** Polls that attempt to predict the results of elections need not only to have a random sample, but also want to include only those people who actually are going to vote. Shortly before a presidential election, a pollster attempted to identify people who actually are going to vote by asking each person in the sample, “Are you planning to vote in the upcoming presidential election?”

---

The National Merit Scholarship Corporation conducts two scholarship programs, National Merit Scholars and National Achievement Scholars. All students may compete in the National Merit program. Only Black students may compete in the National Achievement program. In 2012, 8,064 National Merit scholarships and 791 National Achievement scholarships were awarded. You plan to select a sample of these students for in-depth interviews about their college experience.


**a.** You want a sample of 300 students and plan to take a stratified random sample, stratifying on the program in which the student received a scholarship. How many students would you select from each program if you want the number to be proportional to the number who received a scholarship?
ON YOUR OWN

b. Sometimes survey experts oversample smaller strata by including more people from the smaller strata than they would with proportional sampling. (This allows them to get a good estimate of the parameter from the smaller strata as well as from the larger strata.)

Suppose that you oversample by randomly selecting 100 National Achievement scholars. Of the 200 National Merit scholars in your sample, 8% majored in health sciences. Of the 100 National Achievement Scholars you sampled, 10% majored in health sciences.

i. How many of the National Merit scholars majored in health sciences?
   How many of the National Achievement Scholars majored in health sciences?

ii. What is your estimate of the overall percentage of scholars who majored in the health sciences?

17 Visit the Web site of the American Community Survey at www.census.gov/acs/www/. Then write a report that includes answers to the following questions as well as any other interesting facts you find.

a. Why is this survey taken? What is done with the information collected?

b. What specific questions on the survey would be of interest to students?

c. How is the sample selected? What is the sample size?

18 Visit the Web site of the United States Census Bureau at www.census.gov. Then write a report that includes answers to the following questions as well as any other interesting facts you find.

a. How often is the census taken?

b. When did the census start and why?

c. What is done with the information collected by the census?

d. What questions does the census ask?

e. Is sampling used in the census?

f. How might the census be biased because of nonresponse?

19 Suppose you want to survey your class for some very personal information. If you ask students directly, they might not tell the truth. This would create bias in your survey. Getting honest answers to sensitive questions is a common problem for pollsters. One way to solve this problem is to use a randomized response technique.

Using this technique, the survey designer pairs a sensitive question (to be answered “yes” or “no”) with a harmless question, to which the interviewer could not possibly know the answer and which has a known proportion of “yes” responses. For example, the harmless question might be “Is the coin...
you just (secretly) tossed a head?” When the respondent comes to the sensitive question, he or she secretly rolls a die and flips a coin. If the die lands 1 or 2, then the respondent answers the harmless question about the coin flip. If die lands 3, 4, 5, or 6, then the respondent answers the sensitive question. The interviewer records the number of “yes” answers.

a. Explain why only the respondent knows which question is being answered.

b. Suppose in a sample of 60 people there were 32 “yes” responses. Estimate the proportion of “yes” responses to the sensitive question.

c. Suppose in a sample of 40 people, there were 31 “yes” responses. Estimate the proportion of “yes” responses to the sensitive question.

d. Complete the general statement below showing how the proportion of all “yes” responses is related to the proportion of “yes” responses to a sensitive question and to the proportion of “yes” responses to a harmless question.

Proportion of all “yes” responses $= \ldots$

20 Select a topic that is of interest to students at your school. Write a question about the topic in a way that you think will cause bias. Then rephrase the question in a way that you think will cause bias in a different direction. Give each question to at least 30 different students. Collect responses and compare the results. Did the difference in wording have the effects you thought it would? Explain.

21 The Highline Ice Cream Shop has five flavors of ice cream and 10 different mix-in choices.

a. How many different possible bowls of ice cream are there that contain only one flavor of ice cream and three different mix-ins?

b. How many different possible bowls of ice cream are there that contain two different flavors of ice cream and one mix-in for each flavor? The mix-ins can be the same or different.

c. How many different possible bowls of ice cream are there that contain three different flavors of ice cream (with no mix-in choices)?

d. Which part of this problem is an example of combinations?

22 Find the coordinates of the vertex, the x-intercepts, and the y-intercept of the graphs of each quadratic function.

a. $y = 10x - x^2$

b. $y = x^2 - 8x + 12$

c. $y = (x - 3)^2 - 4$
ON YOUR OWN

23 Solve each equation.
   a. \(3x - 5(x - 3) = 2x - 9\) 
   b. \(x(x - 3) = x^2 - 12\) 
   c. \((2x - 1)^2 = 49\) 
   d. \(2x^2 - 5x - 1 = 2\)

24 Rewrite each expression as a product of linear factors.
   a. \(x^3 + 10x + 24\) 
   b. \(25 - 16x^2\) 
   c. \(x^2 - 12x + 36\) 
   d. \(x^2 + 23x - 50\) 
   e. \(4a^2 - 9b^2\)

25 Write each expression in a simpler equivalent exponential form using only positive exponents.
   a. \(\frac{3x^{-1}}{x^2}\) 
   b. \(\frac{x^2(x^4)^2}{x^3}\) 
   c. \(-6x^5y^3(4x^{-2}y^4)\) 
   d. \(3(21x^2)^{-1}\)

26 Find the indicated angle measure and side lengths.
   a. Find \(m\angle B\).

   b. Find \(AC\).

   c. Find \(BC\) and \(AC\).

   d. Find \(AB\).
Perhaps you have wondered how a pollster can make precise claims about the entire population of the United States based on just a relatively small random sample of people.

For example, a Gallup poll asked Americans if they “think nuclear power plants in the United States are safe or not safe.” Fifty-seven percent of those surveyed said, “safe.” The following technical information about this poll comes from the Gallup Web site.

**Survey Methods**

Results for this Gallup poll are based on telephone interviews conducted March 8–11, 2012, with a random sample of 1,024 adults aged 18 and older, living in the continental U.S., selected using random-digit-dial sampling. For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error is ±4 percentage points. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.

Source: www.gallup.com/poll/153452/americans-favor-nuclear-power-year-fukushima.aspx
**THINK ABOUT THIS SITUATION**

Think about the information about the survey methods used by the Gallup poll.

a. Does this appear to be a trustworthy survey?

b. What sources of possible bias are mentioned?

c. What might be meant by the “margin of sampling error is ±4 percentage points”?

d. What do you think is meant by “with 95% confidence”?

In this lesson, you will develop understanding of the meaning and associated statistical method(s) of each term in the description of the survey methods of the Gallup poll.

**INVESTIGATION 1**

**Variability from Sample to Sample**

Public opinion can be measured with a great deal of accuracy by looking at a relatively small random sample from the population. The first step in understanding how this can be done is to understand how much variability there is among different random samples taken from the same population.

As you work on the problems in this investigation, look for answers to this question:

*How far is the proportion of successes in a random sample likely to be from the proportion of successes in the population from which it was taken?*

1. In Lesson 1, you found the probability of a specified number of successes in a binomial situation. When the sample size \( n \) is large, people usually find it more useful to know the proportion of successes in the sample \( \hat{p} \) rather than the number of successes \( x \). (The symbol \( \hat{p} \) is read “\( p \) hat.”) So, typically, when you hear the results of a survey or poll, the proportion of successes is given. For example, a typical report on the President’s approval rating will say something like, “From our survey of 1,248 adults, we found that 57% approve of the way the President is handling his job.”

a. Suppose that in a sample survey last month, 367 of 1,192 adults approved of the way the President is handling his job. In a similar sample survey this month, 472 of 1,364 adults approved. Was there greater approval for the President among those surveyed this month or among those surveyed last month?

b. The proportion of successes in the sample \( \hat{p} \) is a **point estimate** of the parameter \( p \), the proportion of successes in the population. Write a formula for computing \( \hat{p} \) given the number of successes \( x \) in the sample and the sample size \( n \).
According to the 2010 U.S. Census, about 49% of all 18-year-olds are female.

Suppose that Nicholas does not know this. To estimate the proportion of all 18-year-olds who are female, he will randomly select 100 18-year-olds and count the number of females. In this problem, you will investigate the distribution of all possible outcomes.

a. Nicholas selects his sample and gets 47 females. Compute his point estimate \( \hat{p} \) and locate it with a dot on a copy of the plot below.

b. Nicholas’s estimate from the sample of 0.47 is only 0.02 away from the proportion of successes in the population of 0.49. His sampling error is 0.02, or 2%.

Sampling Error (also, error attributable to sampling or chance error)

Sampling error is the absolute value of the difference between the estimate from a sample and the population parameter. When estimating a population proportion \( p \) from the proportion of successes \( \hat{p} \) in a sample, the sampling error (S.E.) is:

\[
S.E. = |\hat{p} - p|
\]

Usually people can take only one sample. But suppose that Nicholas takes another random sample of 100 18-year-olds and gets 52 females. Add his second point estimate to the plot. What is his sampling error this time?

c. Nicholas continues until he has 1,000 random samples of 100 18-year-olds. The dot plot at the top of the following page shows his 1,000 values of \( \hat{p} \). He has created an approximate sampling distribution. (The exact sampling distribution would have the values of \( \hat{p} \) for all possible random samples.)

i. In how many samples did he get exactly 60 females?

ii. What was the fewest number of females he got in any sample? The largest number?

iii. What is the largest sampling error he got?
d. Suppose that you take a random sample of 100 18-year-olds and compute the proportion who are female.

i. Use the approximate sampling distribution above to estimate the probability that you will get a value of \( \hat{p} \) that is less than 0.39.

ii. Estimate the probability that \( \hat{p} \) will be more than 0.59.

iii. Use the approximate sampling distribution to estimate the probability that you get a sampling error of more than 0.10. Of 0.10 or less.

iv. Make a rough estimate of the probability that you get a sampling error of 0.05 or less.

Nicholas’s approximate sampling distribution illustrates variability in sampling. That is, the point estimate \( \hat{p} \) varies from random sample to random sample. Next you will explore the main factor that affects how large the variability in sampling tends to be.

a. Nicholas notices how spread out the various values of \( \hat{p} \) are in the distribution of Problem 2. Often, the sampling error was large. What do you think that pollsters do to minimize the chance of getting a large sampling error?

b. Nicholas decides to try a larger sample size. Describe how he could make an approximate sampling distribution for samples of size 400 (with \( p \) still equal to 0.49).
c. The dot plot below, an approximate sampling distribution, shows the values of \( \hat{p} \) from 1,000 random samples of size \( n = 400 \).

i. How many 18-year-olds were sampled to get one dot?

ii. What was the largest sampling error this time?

\[
\begin{align*}
0.40 & \quad 0.42 & \quad 0.44 & \quad 0.46 & \quad 0.48 & \quad 0.50 & \quad 0.52 & \quad 0.54 & \quad 0.56 & \quad 0.58 & \quad 0.60 \\
\end{align*}
\]

Proportion of Females

<table>
<thead>
<tr>
<th>Proportion of Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.56</td>
</tr>
<tr>
<td>0.58</td>
</tr>
<tr>
<td>0.60</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\end{align*}
\]

d. Use the approximate sampling distribution to estimate the probability of getting a sampling error of more than 0.10. Of 0.10 or less.

e. Use the approximate sampling distribution to estimate the probability of getting a sampling error of 0.05 or less.

f. By increasing the sample size from \( n = 100 \) to \( n = 400 \), how does the probability of getting a sampling error of 0.05 or less change?

4 A grade of 3 or better typically is considered “passing” on an Advanced Placement (AP) exam. According to the AP Report to the Nation from the College Board, nationwide almost 20% of 2012 high school graduates passed an AP exam during high school. *(Source: The 9th Annual AP Report to the Nation, February 13, 2013, page 37)*

Rose does not know this percentage. To estimate it, she will take a random sample of 100 graduates from the class of 2012.

a. In her sample of 100 graduates, Rose finds that 26 have passed an AP exam. What is her point estimate, \( \hat{p} \)? What is her error attributable to sampling?

b. The dot plot at the top of the following page shows the values of \( \hat{p} \) from 500 different random samples of 100 graduates each. Each sample was taken from the class of 2012 where 20% passed an AP exam during high school. What is the name for this type of dot plot? Locate Rose’s point estimate on the plot.
c. Using the plot, estimate the proportion of random samples that have a sampling error of 0.10 or less. Estimate the proportion of random samples that have a sampling error of 0.05 or less.

d. Does the sampling error for this situation, with \( n = 100 \) and \( p = 0.20 \), tend to be smaller or larger than the one in Nicholas’s approximate sampling distribution in Problem 2 for samples of size 100 and \( p = 0.49 \)?

5 As you saw in the previous problems, the size of the sampling error depends on the sample size \( n \). It also depends on the proportion \( p \) of successes in the population \( p \). Bigger sample sizes make for smaller errors attributable to sampling. Values of \( p \) that are farther from 0.5 make for smaller errors attributable to sampling than values closer to 0.5. The following rule tells you how far apart \( p \) and \( \hat{p} \) are likely to be.

**Rule for Size of Sampling Error**

If you take a random sample of size \( n \) from a population with proportion of successes \( p \), there is a 95% chance that \( p \) and \( \hat{p} \) will be no more than \( 2\sqrt{\frac{p(1-p)}{n}} \) apart. Alternatively, there is a 5% chance that the sampling error will be greater than \( 2\sqrt{\frac{p(1-p)}{n}} \).

a. Compute the value of \( 2\sqrt{\frac{p(1-p)}{n}} \) for the situation when \( n = 100 \) and \( p = 0.20 \). Check whether the rule holds, approximately, for the distribution in Problem 4.

b. Rewrite the first sentence of the rule in the context of Problem 4.

This rule works well only if your sample contains at least 10 successes and at least 10 failures and if the sample size is less than one-tenth of the size of the population.

If the sample size is larger, there is another rule to use, but that is rarely the case in sample surveys.
SUMMARIZE THE MATHEMATICS

In this investigation, you learned to use sampling distributions to estimate the sampling error when taking a random sample from a known population.

a. How would you create an approximate sampling distribution of \( \hat{p} \) for the situation of rolling a pair of dice 50 times and computing the proportion of times they land doubles?

b. What is the sampling error? How is it computed? Can people who conduct sample surveys actually compute the sampling error? Why or why not?

c. If you flip a coin repeatedly and use the proportion of heads in your sample as an estimate of the probability that the coin lands heads, will the sampling error tend to be smaller if you use a sample of 100 flips or a sample of 400 flips? Explain your reasoning.

d. Explain in your own words the meaning of the rule:

\[
\text{If you take a random sample of size } n \text{ from a population with proportion of successes } p, \\
\text{there is a 95% chance that } p \text{ and } \hat{p} \text{ will be no more than } 2 \sqrt{\frac{p(1-p)}{n}} \text{ apart. Alternatively,} \\
\text{there is a 5% chance that the sampling error will be greater than } 2 \sqrt{\frac{p(1-p)}{n}}.
\]

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

CHECK YOUR UNDERSTANDING

According to the 2010 U.S. Census, about 87% of U.S. residents, age 25 and older, are high school graduates. Bernardo does not know this percentage so will take a random sample of 1,500 U.S. residents, age 25 and older to estimate it.

a. What is the value of \( n \)? What is the value of \( p \)?

b. In his sample of 1,500 U.S. residents age 25 and older, Bernardo finds that 1,300 are high school graduates. What is his point estimate, \( \hat{p} \)? What is his sampling error?
c. The dot plot below shows the values of \( \hat{p} \) from 1,000 different random samples, each of 1,500 U.S. residents age 25 and older. What is the name for this type of dot plot? Locate Bernardo’s result on the plot.

![Dot plot showing values of \( \hat{p} \)]

d. From the plot, estimate the proportion of random samples that have a sampling error of 0.03 or less. Estimate the proportion that have a sampling error of 0.02 or less.

e. Does the sampling error here tend to be smaller or larger than the other ones in this investigation? Why is this the case?

f. Compute the value of \( 2 \sqrt{\frac{p(1-p)}{n}} \) for this situation. Write a sentence explaining what the value means in the context of this situation.

---

**INVESTIGATION 2**

**The Margin of Error**

As you worked through the previous investigation, you may have been surprised about how close the proportion \( \hat{p} \) of successes in a random sample tends to be to the proportion \( p \) of successes in the population even when the sample size is as small as 100 out of a population of millions. Of course, larger samples are better, but “large enough” is smaller than most people think.

In Investigation 1, you were able to compute the sampling error \( |p - \hat{p}| \) exactly because you knew the value of both \( \hat{p} \) and \( p \). But pollsters do not know the value of \( p \), or they would not be taking a sample survey.

As you work on the problems in this investigation, look for answers to this question:

**How can you estimate the sampling error when you do not know the proportion \( p \) of successes in the population?**

1. The following rule from the previous investigation will make it possible for you to estimate the sampling error.

   \[
   \text{If you take a random sample of size } n \text{ from a population with proportion of successes } p, \\
   \text{there is a 95\% chance that the difference between } p \text{ and } \hat{p} \text{ will be } 2\sqrt{\frac{p(1-p)}{n}} \text{ or less.}
   \]
Because pollsters do not know the value of $p$, they substitute $\hat{p}$ into the formula in place of $p$:

**Margin of Error**

If you have a random sample of size $n$ from a population with proportion of successes $\rho$, then you can be 95% confident that the difference between $\rho$ and $\hat{p}$ will be

$$2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

or less. The value, $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, is called the margin of error.

- **a.** To see if the substitution of $\hat{p}$ for $p$ makes much difference, compute both $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ and $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for the situation of flipping a fair coin 100 times and getting 47 heads.

- **b.** A survey of 727 randomly selected New Yorkers found that 54% thought that stores should not open on Thanksgiving night. The survey gives a margin of error of about 4%. ([Source: siena.edu/uploadedfiles/home/parents_and_community/community_page/sri/independent_research/Hday1112%20Release_Final.pdf])
  i. Verify this margin of error.
  ii. Interpret this margin of error by filling in the blanks in this sentence:

    If all New Yorkers could have been asked if stores should open on Thanksgiving night, you are 95% confident that the percentage who would have said ________ is no farther than ________ away from the estimate from the sample of ________.

- **c.** In national polls, major polling organizations typically use a sample size of about 1,200. Suppose that a poll finds that 45% of the 1,200 U.S. adults surveyed approve of the job the President is doing. Compute and interpret the margin of error for this poll.

- **d.** Suppose that Sample X is larger than Sample Y, but the proportion of successes in the samples are the same. How do the margins of error differ? Explain your answer.

- **e.** Suppose that the proportion of successes in Sample P is the same as that in Sample Q. However, the sample size for Sample P is four times as large as that for Sample Q. What is the relationship between their margins of error?

**2** The margin of error tells you the farthest apart $p$ and $\hat{p}$ are likely to be. In 95% of all random samples, the difference between $p$ and $\hat{p}$ will be no larger than the margin of error. But that means that in 5% of all random samples, $p$ and $\hat{p}$ are farther apart than the margin of error. Unfortunately, we usually do not know when that is the case. One time that we do is for polls taken right before an election, when the vote on election day tells us how well predictions from polling turned out. In the presidential election of November 2008, Barack Obama got 52.9% of the vote. The table on the next page gives the results from 15 polls taken right before election day. ([Source: www.realclearpolitics.com/epolls/2008/president/national.html])

---

**LESSON 3 | Margin of Error: From Sample to Population**
<table>
<thead>
<tr>
<th>Name of Poll</th>
<th>Number of Likely Voters Surveyed</th>
<th>Margin of Error</th>
<th>Percent Saying They Would Vote for Obama</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marist</td>
<td>804</td>
<td>4.0</td>
<td>52</td>
</tr>
<tr>
<td>Battleground (Lake)</td>
<td>800</td>
<td>3.5</td>
<td>52</td>
</tr>
<tr>
<td>Battleground (Tarrance)</td>
<td>800</td>
<td>3.5</td>
<td>50</td>
</tr>
<tr>
<td>Rasmussen Reports</td>
<td>3,000</td>
<td>2.0</td>
<td>52</td>
</tr>
<tr>
<td>Reuters/C-SPAN/Zogby</td>
<td>1,201</td>
<td>2.9</td>
<td>54</td>
</tr>
<tr>
<td>IBD/TIPP</td>
<td>981</td>
<td>3.2</td>
<td>52</td>
</tr>
<tr>
<td>FOX News</td>
<td>971</td>
<td>3.0</td>
<td>50</td>
</tr>
<tr>
<td>NBC News/Wall St. Journal</td>
<td>1,011</td>
<td>3.1</td>
<td>51</td>
</tr>
<tr>
<td>Gallup</td>
<td>2,472</td>
<td>2.0</td>
<td>55</td>
</tr>
<tr>
<td>Diageo/Hotline</td>
<td>887</td>
<td>3.3</td>
<td>50</td>
</tr>
<tr>
<td>CBS News</td>
<td>714</td>
<td>—</td>
<td>51</td>
</tr>
<tr>
<td>ABC News/Wash Post</td>
<td>2,470</td>
<td>2.5</td>
<td>53</td>
</tr>
<tr>
<td>Ipsos/McClatchy</td>
<td>760</td>
<td>3.6</td>
<td>53</td>
</tr>
<tr>
<td>CNN/Opinion Research</td>
<td>714</td>
<td>3.5</td>
<td>53</td>
</tr>
<tr>
<td>Pew Research</td>
<td>2,587</td>
<td>2.0</td>
<td>52</td>
</tr>
</tbody>
</table>

d. Verify the margin of error given for the Battleground (Lake) poll.

g. What was the sampling error for the Battleground (Lake) poll?

c. For the Battleground (Lake) poll, was it the case that the difference between \( p \) and \( \hat{p} \) was less than or equal to the margin of error?

d. Supply the margin of error for the CBS News poll.

e. For which poll(s) was the difference between \( p \) and \( \hat{p} \) larger than the margin of error?

f. Suppose that you examine 100 different political polls. What is the expected number of polls where the sampling error is larger than the margin of error?

g. If you examine 15 political polls, what is the expected number where the sampling error is larger than the margin of error? So, is there any reason to be concerned by the result of Part e? Explain.

3 Which two of the following are true statements about the margin of error?

A. In 95% of all random samples, the absolute value of the difference between \( p \) and \( \hat{p} \) will be less than or equal to the margin of error.

B. If you take a random sample, the absolute value of the difference between \( p \) and \( \hat{p} \) will be less than or equal to the margin of error.

C. The margin of error tells you exactly how far apart \( p \) and \( \hat{p} \) are.

D. The margin of error depends only on the sample size.

E. If you want a smaller margin of error, you should take a larger random sample.

F. In a sample survey, the margin of error is equal to the sampling error.
Spin a penny 20 times by placing it on edge on the floor and flicking with your finger. Let the penny spin freely until it lands heads up or heads down. Count the number of heads.

4

a. Do you think that spinning a penny is fair? Explain your reasoning.

b. Combine results with the rest of your class until you have a total of 400 spins. Compute \( \hat{p} \), your estimate of the probability a spun penny lands heads up. Test the statistical significance of your result (compared to a fair coin) by computing the \( P \)-value.

c. How can you use the margin of error rather than a \( P \)-value to help you decide whether you think that spinning a penny is fair?

### SUMMARIZE THE MATHEMATICS

In this investigation, you learned how to compute and interpret the margin of error connected with the estimate of the proportion of successes in a population.

a. What does it mean if you are told that the margin of error for a survey is 3%?

b. What is the formula for computing the margin of error? When can you use this formula?

c. What is the difference between sampling error and margin of error?

d. Read the following definition of margin of error from a newspaper. Do you think it is a good explanation for the average person? How would you improve it?

"Margin of sampling error: Political polls can never survey all voters. Pollsters use mathematical formulas to determine how much error might be in their results. The more people surveyed, the smaller the margin will be."


Be prepared to explain your ideas to the class.
The first political poll in the United States was conducted by the *Harrisburg Pennsylvanian* before the 1824 presidential election. The newspaper polled 532 voters in Wilmington, Delaware. Three hundred thirty-five of those polled said they preferred Andrew Jackson.

**a.** What proportion of those polled preferred Andrew Jackson?

**b.** What must be true about the method of sampling so that you can compute the margin of error? Is this likely to be the case? Explain your thinking.

**c.** Regardless of your answer to Part b, compute the margin of error for this poll.

**d.** What sample size would have cut the margin of error in half? Justify your answer.

### Interpreting a Confidence Interval

Sometimes news stories, especially those in the medical field, will report a confidence interval rather than a margin of error. For example, confidence intervals were reported for a research study that followed 770 people who were newly diagnosed with type II diabetes and had no diabetic retinopathy (damage to the eye caused by diabetes). After six years, 90 of them had diabetic retinopathy. A 95% confidence interval for the proportion of all such diabetics with diabetic retinopathy can be given in this form: (9.4, 14.0)%.


As you work on the problems in this investigation, look for answers to this question:

How is a 95% confidence interval computed and how should it be interpreted?

1 Refer to the study about diabetic retinopathy above.

   **a.** Compute \( \hat{p} \), the point estimate, for the proportion of all newly diagnosed type II diabetics who would develop diabetic retinopathy within six years. Write a newspaper headline using this point estimate.

   **b.** Compute the margin of error.

   **c.** Examine the 95% confidence interval, (9.4, 14.0)%. Tell how each of the numbers was computed.
d. The endpoints of a 95% confidence interval (sometimes abbreviated 95% CI) are given by this formula:

\[ \hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]

i. What is the name for the first term of this formula? What is the name for the last term of this formula?

ii. How would you write the 95% CI using interval notation?

e. Here is how a 95% confidence interval should be interpreted.

**Interpretation of a 95% Confidence Interval**

With random sampling, we are 95% confident that the proportion \( p \) of successes in the population lies within the confidence interval.

Use the sentence above as a guide to interpret the confidence interval about diabetic retinopathy in the context of the situation.

2 Answer the following questions for each situation below:

**Question 1:** What is the population?

**Question 2:** What parameter of the population is the survey trying to estimate?

**Question 3:** What is the point estimate of that parameter?

**Question 4:** What is the 95% confidence interval for that parameter?

**Question 5:** What is the interpretation, in context, of this interval?

a. A Gallup poll found that 96% of Americans had a favorable view of Canada; more than for any other country. According to Gallup, the poll was conducted using telephone interviews with a random sample of 1,029 adults living in the United States.

   *(Source: www.gallup.com/poll/152735/americans-give-record-high-ratings-several-allies.aspx)*

b. A survey of the parents of American teenagers who use the Internet found that two-thirds of the parents said they were active on social networking sites themselves. This poll had a margin of error of 4.5%.

   *(Source: www.pewinternet.org/Reports/2012/Teens-and-Privacy.aspx)*

3 To estimate the prevalence of alcohol consumption during pregnancy, a stratified random sample of 12,611 mothers from Maryland who delivered live infants during the years 2001–2008 filled out a questionnaire. The results included this statement, “Nearly 8% (95% confidence interval 7.1–8.4) of mothers from Maryland reported alcohol consumption during the last 3 months of pregnancy.” *(Source: Cheng, Diana; Kettinger, Laurie; Uduhiri, Kelechi; Hurt, Lee, “Alcohol Consumption During Pregnancy: Prevalence and Provider Assessment.” Obstetrics & Gynecology. 117 (February 2011) pages 212–217)*

   a. The actual percentage from the sample was 7.75%. Use this percentage to compute a 95% confidence interval for this situation.
b. Large surveys like this one often use stratified random samples. The formula for a confidence interval for a stratified random sample is a bit different from the formula you are using. Thus, your computation may not always exactly match that reported in the media, but it should be close. Is there much difference between the confidence interval reported in the research article and the one you computed in Part a?

4You can say that you are 95% confident that the true population proportion lies in your 95% confidence interval. But what, exactly, does the phrase “95% confident” mean? Suppose you use the method of constructing the confidence interval with many different random samples. Then, for some samples, the true population proportion will be inside the confidence interval and for some samples it will not. But, overall, you expect the interval to “capture” the true proportion of successes 95% of the time. The following activity illustrates exactly what this means.

a. On your own, use a calculator or software such as TCMS-Tools Simulation software to generate 50 random digits. Count the number of digits that are even. (Remember that 0 is an even digit.) Check your count by also counting the number that are odd.

b. Using your random sample, find a 95% confidence interval for the percentage of all random digits that are odd.

c. Draw a heavy line segment to represent your confidence interval along one of the vertical lines in a copy of the chart below. For example, suppose you are the fifth student to add your confidence interval and you get 24 even digits in your sample of 50, for a sample proportion of 0.48. Then your 95% confidence interval is $0.48 \pm 0.14$, or $(0.34, 0.62)$. You would draw a line segment from 0.34 to 0.62 along the vertical line above the 5 on the “Student” axis.

95% Confidence Intervals for the Proportion of Even Digits

![Chart](chart.png)

d. Are all intervals from the students in your class exactly the same? Should they be?

e. What is the true proportion of random digits that are even? What proportion of the intervals from your class “captured” this true proportion? Is this about what you would expect?
**SUMMARIZE THE MATHEMATICS**

In this investigation, you learned how to compute and interpret a 95% confidence interval.

a. What is the relationship between a margin of error and a 95% confidence interval?

b. Is \( \hat{p} \) always in the confidence interval? Is \( p \) always in the confidence interval? Explain.

c. Explain the meaning of “95% confident.”

d. Why is it better for pollsters to report a confidence interval (or margin of error) rather than giving just a point estimate?

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

**CHECK YOUR UNDERSTANDING**

Suppose you have designed a map that describes a walking tour of your neighborhood. To determine whether it is worth marketing, you would like to know what percentage of people in your neighborhood would be interested in buying such a map. So, you survey 100 residents of your neighborhood, selected at random. Thirty-four people say they would be interested in buying such a map.

a. What is your population?

b. What percentage are you trying to estimate?

c. What is your point estimate of that percentage?

d. What is the 95% confidence interval for that percentage?

e. Give an interpretation of your 95% confidence interval in the context of this situation.

f. Explain what you mean when you say that you are “95% confident.”
ON YOUR OWN

APPLICATIONS

1. In the United States, 27% of households with mortgages owe more than their house is worth. This is called being “underwater.” Jesse does not know this percentage and picks a random sample of 300 U.S. households with mortgages to estimate the percentage underwater. (Source: money.cnn.com/2011/02/09/real_estate/underwater_mortgages_rising)

   a. What is the value of \( n \)? What is the value of \( p \)?

   b. In Jesse’s random sample of 300 households, 92 are underwater. What is Jesse’s point estimate \( \hat{p} \) of the proportion of all U.S. households that are underwater? What is his sampling error?

   c. The dot plot below shows the values of \( \hat{p} \) from 1,000 different random samples, each of 300 U.S. households. What is the name for this type of dot plot? Locate Jesse’s result on the plot.

   ![Dot Plot Image]

   d. From the plot, estimate the proportion of random samples that have a sampling error of 0.05 or less.

   e. Compute the value of \( 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \) for this situation. Write a sentence explaining what the value means in the context of this situation.

2. Over a six-month period, only twelve percent of Alaska Airlines flights did not arrive on time. Jasmine does not know this percentage and so takes a random sample of 100 departing flights to try to estimate it. (Source: www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/press_releases/2012/dot089_12/html/dot089_12.html)

   a. What is the value of \( n \)? What is the value of \( p \)?

   b. In Jasmine's random sample of 100 flights, 20 were late. What is Jasmine's point estimate \( \hat{p} \) of the proportion of all Alaska Airlines flights that are late? What is her sampling error?
c. The approximate sampling distribution below shows the values of \( \hat{p} \) from 1,000 different random samples, each of 100 departing flights. From the plot, estimate the proportion of times that a random sample of size 100 will give a proportion of flights that are late as high or even higher than Jasmine got in her sample.

![Sampling distribution graph]

\[
\text{Proportion Late} \quad 0.00 \quad 0.05 \quad 0.10 \quad 0.15 \quad 0.20 \quad 0.25
\]

\[
\begin{array}{c}
0.25 \\
0.20 \\
0.15 \\
0.10 \\
0.05 \\
0.00 \\
\end{array}
\]

\[ \hat{p} \]

\[ n \]

\[
\sqrt{n}
\]

\[ p \]

\[ 1 - p \]

\[ \text{Margin of Error: From Sample to Population} \]

\[ \text{LESSON 3} \]

\[ 415 \]

\[ \text{Program: MMH Core Plus Math Component: TCMS-SE Vendor: Six Red Marbles Grade: 12 Final Pass} \]

d. From the plot, estimate the proportion of random samples that have a sampling error of 0.05 or less.

e. Compute the value of \( 2\sqrt{\frac{p(1-p)}{n}} \) for this situation. Write a sentence explaining what the value means in this context.
3 According to the Australian paper, the Daily Telegraph, “A Galaxy Research survey of more than 1,200 people aged 18–39 has found 47 per cent have chosen to end it with their partner on or around Valentine’s Day after taking stock of their relationship.” (Source: www.dailytelegraph.com.au/lifestyle/valentines-day-a-popular-time-to-break-up/story-e6frf0oi226005579514)

   a. How many people in the sample of 1,200 people have broken up with someone around Valentine’s Day?

   b. What is the parameter that Galaxy Research is trying to estimate?

   c. What must be true about the method of sampling so that you can compute the margin of error?

   d. Assuming that the method in Part c was used, compute the margin of error for this poll.

   e. Write an interpretation of the margin of error, in context.

4 Suppose you place a penny on the ground at your school and observe whether the next student to walk by picks it up. After doing this 75 times, you find that 30 students picked up the penny. Assume that the 75 students in your sample can be considered a random sample of students from your school.

   a. What parameter are you trying to estimate?

   b. What is your point estimate of the proportion of all students at your school who would pick up the penny?

   c. Compute the margin of error for this survey.

   d. Write an interpretation of the margin of error, in context.

5 A survey of the ethics of high school students reported that 11,781 out of 22,912 students said they had cheated on one or more tests in the last year. (Yet, 96% said that it was important for people to trust them!) (Source: charactercounts.org/pdf/reportcard/2012/ReportCard-2012-DataTables-HonestyIntegrityCheating.pdf)

   a. What parameter about cheating was the survey trying to estimate?

   b. What is the point estimate of that parameter?

   c. The researchers did not select the sample randomly, but attempted to acquire responses from a wide variety of high schools. Nevertheless, compute a margin of error.

   d. The survey reported, “The survey findings have an error margin of plus or minus less than one percent” Do you agree with their margin of error? Write a more complete sentence reporting the margin of error.
6 At the beginning of this lesson (page 399), a Gallup poll is reported that asked Americans if they “think nuclear power plants in the United States are safe or not safe.” Fifty-seven percent of the 1,024 surveyed said, “safe.”

a. What parameter is Gallup trying to estimate?

b. The margin of error is given as ±4%. Is this close to that given by your formula?

c. Explain what is meant by “sampling error.”

d. If Gallup had reported a 95% confidence interval instead of the margin of error for this situation, what would that interval be?

e. Interpret this interval in the context of this situation. Then, explain what is meant by “95% confidence.”

7 A study was conducted of children being cared for by a relative (called *kinship care*) after being removed from their family because of maltreatment. Of the 572 children in kinship care, 173 had behavioral problems. These children were not randomly selected, but the sample was selected to be “nationally representative.” (Source: www.medpagetoday.com/Pediatrics/GeneralPediatrics/24743 based on Sakai S, et al. “Health outcomes and family services.” *Arch Pediatr Adolesc Med* 2011; 165(2): 159–165.)

a. What is the population here?

b. What parameter were the researchers trying to estimate? What is the point estimate of that parameter?

c. What is the 95% confidence interval for that parameter? What assumption must you make?

d. What is the interpretation, in context, of this interval?

8 The approximate sampling distributions in Investigation 1 were made using simulation. You also can make exact sampling distributions using the binomial probability formula developed in Lesson 1. In this task, you will construct an exact sampling distribution for the situation of flipping a fair coin 20 times and computing the proportion of times it lands heads. Before beginning, scan the parts of this task to help you make strategic decisions about technology tools that may be helpful in your work.

a. Use the binomial probability formula to compute the probability that the proportion of heads will be 0.6.
ON YOUR OWN

b. Use the `binompdf(n, p, x)` function to compute probabilities for the remaining possible proportions of heads.

c. Make a plot with proportion of heads on the horizontal axis and probability on the vertical axis.

d. Suppose you flip a coin 20 times and get 14 heads. What is the sampling error in your estimate of the probability that a coin will land heads? Use your exact sampling distribution to find the probability of getting a sampling error no larger than that.

9 Examine the formula for the margin of error, \( E = 2\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \).

a. Suppose you thought about taking a sample of size 100, but find that the margin of error is three times as large as you would like. What sample size should you use? Justify your answer.

b. Suppose you thought about taking a sample of size 100, but find that the margin of error is four times as large as you would like. What sample size should you use? Justify your answer.

c. In general, by what factor should you increase the sample size if you want to cut your margin of error by half? By \( \frac{1}{n} \), where \( n \) is an integer greater than 1?

10 Suppose that you expect to get a sample proportion around 0.5 so that the margin of error \( E \) is

\[
E = 2\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 2\sqrt{\frac{0.5(1 - 0.5)}{n}}.
\]

a. Simplify the expression on the right.

b. Make a graph of the function with sample size \( n \), from 1 to 1,000, on the horizontal axis and margin of error on the vertical axis.

c. Describe how the margin of error changes as the sample size increases.

11 Suppose that you have a sample size of \( n = 100 \).

a. Consider \( E = 2\sqrt{\frac{\hat{p}(1 - \hat{p})}{100}} \), where \( E \) is the margin of error. Make a graph of this function, placing \( \hat{p} \) on the horizontal axis and \( E \) on the vertical axis.

b. Which value of \( \hat{p} \) gives the largest margin of error?

c. By how much does the margin of error change with a change in \( \hat{p} \) from 0.5 to 0.6? From 0.8 to 0.9?
The dot plots in Investigation 1 were created using simulation software. Below is a screen from three simulations of a binomial distribution with $n = 100$ and $p = 0.5$.

**a.** How many random samples were generated for each of the three simulations?

**b.** Explain why the three graphs are not identical.

**c.** Notice that the labels on the horizontal axis are the number of successes rather than the proportion of successes. How would the shapes of these distributions change if the proportion of successes had been graphed?

**d.** What are the sampling errors for the 158th sample for each simulation?

The following question appeared on the 10th-grade assessment for the state of Wisconsin. (Source: oea.dpi.wi.gov/files/oea/pdf/mathrelease10.pdf)

Three candidates are running for mayor in the town of Morganville. The local newspaper conducted two surveys asking potential voters for whom they plan to vote. Survey 1 was taken six months before the election and Survey 2 was taken three months before the election. The results of the two surveys are shown below.
ON YOUR OWN

Based on the results of both surveys, which of these statements is true?

Statement 1: Crawford will definitely win the election.
Statement 2: More people will vote for Baxter than Adams in the election.
Statement 3: Crawford is spending the most money of the three candidates.
Statement 4: Baxter gained voter support between Survey 1 and Survey 2.

a. What answer do you think was scored as correct?
b. Why is the question actually impossible to answer? That is, what crucial information is not given that would help you decide whether any of the four statements is necessarily true?

A trustworthy sample survey is unbiased (has no systematic tendency to get an estimate that is too large or too small, on average). It also has a small margin of error.

a. How do you design a survey to minimize bias in sampling?
b. How do you design a survey to get more precision?

Jeri said that confidence intervals were difficult for her to understand at first. She said it took her a while to realize that what you are 95% confident in is not really a particular interval, but rather a method of generating intervals. Explain what you think she means by this statement.

Sometimes confidence intervals (or margins or error) are computed for proportions for two related populations. These then are used to compare the populations.

a. A report on the 2011 mayoral race in Chicago had the following information. Verify that the polls do “overlap.” Then explain what is meant by the last sentence of this report.

Another Day, Another Debate, Another Poll
When ABC-7 polled 600 people, Rahm Emanuel stood at 54 percent. Now, the Chicago Tribune and WGN have polled 718 people, and it shows Emanuel sitting at 49 percent. Technically, the margin of error means these polls overlap.

Source: www.nbcchicago.com/blogs/ward-room/Race-to-Run-Off-115886274.html

b. The study in Applications Task 7 also examined 736 children placed in foster care, of which 391 had behavioral problems. Compute and interpret a 95% confidence interval for this situation. Are you convinced that the proportion of all children placed in kinship care and the proportion of all children placed in foster care who have behavioral problems are different? Explain.
ON YOUR OWN

17 Your calculator will compute confidence intervals when given just the sample size \( n \), the number of successes in the sample \( x \), and the level of confidence desired, which typically is 95%. Find the function \( 1\text{-PropZInt} \) in the TESTS menu. After entering the number of successes \( x \) and the sample size \( n \), enter 0.95 for a 95% confidence level (C-Level), highlight Calculate and press Enter. The calculator uses a factor of 1.96 rather than 2 in the formula for the margin of error, which can make a slight difference in the interval endpoints.

a. Use your calculator to compute a 95% confidence interval for the situation in Applications Task 7 and compare it to the interval computed using the formula for the margin of error.

b. Use your calculator to compute a 95% confidence interval for the situation in Applications Task 6. You must input \( x \) the number of successes, as a whole number, so compute that first.

c. How can you use the \( 1\text{-PropZInt} \) command to find the margin of error? Use your method to determine the margin of error for the situation in Applications Task 6.

d. Use your calculator to determine the margin of error for this situation: Fifty volunteers were asked by the Massachusetts Public Interest Research Group to phone their credit-card company and ask for a lower interest rate, saying they would switch to another company unless given one. Fifty-six percent of them got a lower rate within 5 minutes. (Source: uspirg.org/sites/pirg/files/reports/Deflate_Your_Rate_USPIRG.pdf)

18 Refer to Extensions Task 17 about using your calculator to compute confidence intervals. The confidence intervals you have constructed gave 95% confidence. In some situations people want to have more than 95% confidence that the population proportion is in their interval.

a. If you want to have, say, 99% confidence in your interval, should it be wider or narrower than the 95% confidence interval? Explain your reasoning.

b. A Harris Poll conducted online with 2,016 adult Americans found that 21% said that they had some sort of tattoo. (Source: www.harrisinteractive.com/NewsRoom/HarrisPolls/tabid/447/mid/1508/articleid/970/ctl/ReadCustom%20Default/Default.aspx) Compute a 95% confidence interval and a 99% confidence interval for this situation. Compare the widths. Are they consistent with your answer to Part a?

19 The 1948 presidential election is famous because the polling organizations predicted that Dewey would beat Truman. Shortly before the election, Newsweek published a poll of 50 of the nation’s “leading political writers.” All 50 of them predicted a Dewey victory. (Source: dailynightly.msnbc.msn.com/archive/2008/01/09/565536.aspx)

a. What population proportion was Newsweek trying to estimate?
b. What goes wrong when you try to use the formula for the margin of error?

c. The margin of error rule (page 404) says that “This rule works well only if your sample contains at least 10 successes and at least 10 failures.” Is this condition satisfied in the case of the Newsweek poll?

20 In Unit 1, Interpreting Categorical Data, you learned about relative risk. In the study of kinship versus foster care presented in Applications Task 7 and Reflections Task 16 Part b, of the 572 children in kinship care, 173 had behavioral problems. Of the 736 children in foster care, 391 had behavioral problems.

a. Compute and interpret the relative risk for this situation.

b. Computing a 95% confidence interval for relative risk can be complicated, so a calculator may be found at www.medcalc.org/calc/relative_risk.php, for example. First, compute the relative risk. Note that you must input the number with behavioral problems and the number without for each group (not the sample size \( n \)). Does the relative risk agree with the one you computed in Part a?

c. Interpret the confidence interval computed by the online calculator.

d. Compare the relative risk formula used by the online calculator with the representation that you used in Unit 1 (page 9).
21 Write each sum of rational expressions in equivalent form as a single algebraic fraction. Then simplify the result as much as possible.
   a. \( \frac{x}{4} + \frac{3x + 4}{4} \)
   b. \( \frac{3x}{5} + \frac{2x + 1}{2} \)
   c. \( \frac{4x + 5}{x} + 3x \)

22 Tonya has a credit card balance of $2,345. Her annual interest rate on any unpaid balance is 21% compounded monthly. Tonya makes a payment of $400 each month.
   a. How much interest will be added to Tonya’s balance in the first month?
   b. What will her balance be after she makes the first payment?
   c. Making this payment, and assuming that she does not make any new charges, how long will it take Tonya to pay off the entire credit card balance?
   d. How much interest will she have paid by the time she reduces the balance to zero?

23 Solve each inequality and represent the solution using interval notation and a number line.
   a. \( 10 - 8x > 3(6 - 2x) \)
   b. \( -12 + 3(x - 5) \leq -9 - 6x \)
   c. \( (x + 5)(x - 6) > 0 \)
   d. \( -(x + 5)(x - 6) > 0 \)

24 Suppose that a softball is hit from a height of 1.2 meters with an initial upward velocity of 21 meters per second. Recall that the function rule for the height of a ball is \( h(t) = -4.9t^2 + v_0t + h_0 \), where \( v_0 \) is the initial upward velocity in meters per second and \( h_0 \) is the initial height in meters.
   a. Write a function rule that will give the height of the softball after any number of seconds.
   b. How high will the ball be after 1 second?
   c. At what times will the ball be more than 18 meters above the ground?
   d. If the ball is not caught, when will it hit the ground?

25 The table below shows the types of cell phones owned by random samples of U.S. adults who were surveyed in May and December of 2011. (Source: libraries.pewinternet.org/files/2012/04/Topline_for_-e_reading_report_4_surveys.pdf)

<table>
<thead>
<tr>
<th></th>
<th>May 2011</th>
<th>December 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone</td>
<td>888</td>
<td>1,404</td>
</tr>
<tr>
<td>Basic Cell Phone</td>
<td>1,025</td>
<td>1,194</td>
</tr>
<tr>
<td>No Cell Phone</td>
<td>364</td>
<td>388</td>
</tr>
<tr>
<td>Total</td>
<td>2,277</td>
<td>2,986</td>
</tr>
</tbody>
</table>

Lesson 3 | Margin of Error: From Sample to Population
ON YOUR OWN

a. For each survey, what percent of the people surveyed owned a smartphone?
b. Perform a chi-square test of homogeneity and state your conclusions.

26 Rewrite each expression as a product of linear factors.
   a. $a^2 + 7a + 12$
   b. $3x^2 - 13x - 10$
   c. $4s^2 + 12s + 9$
   d. $t^3 - 8t^2 - 20t$
   e. $30x^2 + 3x - 9$

27 Solve each quadratic equation.
   a. $x^2 = x + 2$
   b. $2x^2 + x - 1 = 0$
   c. $2x^2 - 5x = 3$
   d. $2x^2 + 2 = 5x$
   e. $6x^2 + 5x + 1 = 0$

28 Destiny deposited $1,000 into a savings account that earns 3% interest compounded continuously. Assume that she does not withdraw any money from the account.
   a. Write a formula that could be used to find the balance after $t$ years.
   b. How much money will be in the account after:
      i. six months?
      ii. one year?
      iii. two years?
   c. How much interest will she have earned after 10 years?
   d. How long will it take for her money to double?

29 As a fundraiser, the senior class at a high school in Holland, Michigan is considering selling class sweatshirts. They first need to analyze their costs. The printing company will charge $150 to set up the artwork and $8 per sweatshirt.
   a. Write a formula for the average cost of a sweatshirt if the class purchases $n$ sweatshirts.
   b. How many sweatshirts do they need to purchase so that the average cost per sweatshirt is less than $10?c. Identify any horizontal or vertical asymptotes of the graph of the average cost function. Then explain their meaning in terms of the context of the situation.
30 For each graph, write a function rule that produces the graph.

a. 

b. 

c. 

d. 

31 Consider the rectangular prism and square pyramid shown below.

For each shape, complete the following.

a. Without tracing, make a careful sketch of the polyhedron.

b. How many symmetry planes does the shape have? Describe or draw each one.

c. How many axes of symmetry does each shape have? Describe or draw each one and then identify the angles of rotation associated with each axis of symmetry.

d. Imagine a plane that intersects the shape and is parallel to its base. Is the polygon formed by the intersection of the plane and the polyhedron congruent to, similar to, or neither congruent to nor similar to the base of the polyhedron?
Binomial situations often arise in surveys and experiments. In this unit, you have developed and studied statistical methods for making sense of binomial situations. In the first lesson, you reviewed rules of probability, developed the binomial probability formula to compute binomial probabilities exactly, and used the distribution of the proportion of successes to find P-values and determine statistical significance. In the second lesson, you learned the characteristics of a well-designed survey and how to identify sources of bias in the method of selecting a sample and in the method of obtaining the response. In the third lesson, you learned what is meant by a margin of error and a confidence interval and how a survey can yield an estimate of the population parameter that is reasonably precise even though only a small portion of the population was in the sample. The tasks in this final lesson will help you review and synthesize these important ideas.

1 In the United States, 8.6% of high school students (in grades 9–12) attend a church-related school. (Source: www.census.gov/compendia/statab/2011/tables/11s0234.pdf)

a. Suppose that you pick 2 high school students at random. Demonstrate how to use the Multiplication Rule to compute the probability that neither of them attends a church-related school.

b. Suppose that you select 10 high school students at random. Use the binomial probability formula to compute the probability that exactly 2 of them attend a church-related school.

c. To see if the percentage is larger in the northeastern United States, Carlos takes a random sample of 200 high school students from just the northeast. Twenty-seven of them attend a church-related school. What is the probability of getting 27 or more students who attend a church-related school out of a sample of 200 students if, overall, 8.6% of students in the northeast attend a church-related school? What is the name for the probability you just computed? Is the result from the sample statistically significant? Explain what is meant by “statistically significant.”
2 A Gallup poll estimated that the unemployment rate in the U.S. was 9.8%. Read the following excerpts from their explanation of how the poll was conducted.

Survey Methods

Results are based on telephone interviews conducted as part of Gallup Daily tracking Jan. 2 to 31, 2011, with a random sample of 18,778 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia, selected using random-digit-dial sampling.

For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error is ±0.7 percentage points.

In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.


a. Describe the parameter being estimated. What is the estimate?

b. The Gallup poll tries to make sure that all regions of the country are proportionally represented. At the time of the survey, about 37% of the population of the United States lived in the South. How many people in the South should have been in the sample?

c. Refer to the questions in Problem 3 of Lesson 2, Investigation 1 (page 378). Answer the questions that can be answered based on the information in the article.

d. Is there any indication that the method of conducting this poll might result in an untrustworthy estimate? What else might you want to know in order to decide?

e. What margin of error is reported? What is the 95% confidence interval?

f. Use the formula to compute the margin of error. Does your computation match the margin of error reported?

g. Can you compute the sampling error in this case? If so, do it. If not, explain why not. If there is sampling error, does that mean that a mistake has been made in the survey?

3 A recent survey of 1,033 American adults found that 53 percent of those polled said that chocolate chip cookies are their favorite kind of cookie. The survey was conducted by a reputable firm, so you may assume the sample was selected at random. (Source: Matthew Reynolds, “Chocolate Chip Remains Cookie King” Modern Baking, Vol. 23, No. 13 (October 2009))

a. What is the population?

b. What percentage was the survey trying to estimate?

c. What is their point estimate of that percentage?

d. Compute both the margin of error and the 95% confidence interval.
e. Give an interpretation of the 95% confidence interval in the context of this situation.

f. Why is it important that the sample be selected at random?

g. Suppose that the true percentage of all American adults who would say that chocolate chip cookies are their favorite is 52%. What is the sampling error in this case?

**SUMMARIZE THE MATHEMATICS**

Many questions about probability can be resolved using the techniques you have learned for analyzing binomial situations.

a. What are the Multiplication Rule for Independent Events and the Addition Rule for Mutually Exclusive Events? When can each be used?

b. Describe how to compute the probability of getting $x$ or fewer successes in a binomial situation with $n$ trials and probability of success $p$ using:
   i. the binomial probability formula.
   ii. a binomial probability function on your calculator or statistical software.

c. Describe the reasoning behind the binomial probability formula.

d. How do you compute a $P$-value? How can you tell whether a result is statistically significant? What does it mean if a result is statistically significant?

e. What are the characteristics of a trustworthy survey?

f. What is bias in sampling? What is response bias?

g. What is the sampling error and how is it different from the margin of error? Describe the method of computing each of them.

h. What is it that you are 95% confident is in the confidence interval? What is the meaning of “95% confident”?

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

**CHECK YOUR UNDERSTANDING**

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