6.4 Select and Draw Conclusions from Samples

Before	You used statistics to describe sets of data.
Now	You will study different sampling methods for collecting data.
Why?	So you can interpret the results of a survey, as in Ex. 27.

Key Vocabulary
- population
- sample
- unbiased sample
- biased sample
- margin of error

COMMON CORE

CC.9-12.S.IC.1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population.*

READING
A survey of an entire population is called a *census*.

A **population** is a group of people or objects that you want information about. When it is too difficult, time-consuming, or expensive to survey everyone in a population, information is gathered from a **sample**, or subset, of the population. Some methods for selecting a sample are described below.

In a *self-selected sample*, members of a population can volunteer to be in the sample.

In a *systematic sample*, a rule is used to select members of a population, such as selecting every other person.

In a *convenience sample*, easy-to-reach members of a population are selected, such as those in the first row.

In a *random sample*, each member of a population has an equal chance of being selected.

EXAMPLE 1 Classify samples

BASEBALL A sportswriter wants to survey college baseball coaches about whether they think wooden bats should be mandatory throughout college baseball. Identify the type of sample described.

  a. The sportswriter contacts only the coaches that he has cell phone numbers for in order to get quick responses.

  b. The sportswriter mails out surveys to all the coaches and uses only the surveys that are returned.

Solution

  a. The sportswriter selected coaches that are easily accessible. So, the sample is a convenience sample.

  b. The coaches can choose whether or not to respond. So, the sample is a self-selected sample.

BIAS IN SAMPLING In order to draw accurate conclusions about a population from a sample, you should select an *unbiased sample*. An **unbiased sample** is representative of the population you want information about. A sample that overrepresents or underrepresents part of the population is a **biased sample**.

EXAMPLE 2   Identify a biased sample

CONCERT ATTENDANCE The manager of a concert hall wants to know how often people in the community attend concerts. The manager asks 50 people standing in line for a rock concert how many concerts per year they attend. Tell whether the sample is *biased* or *unbiased*. Explain your reasoning.

Solution

The sample is biased because people standing in line for a rock concert are more likely to attend concerts than people in general.

CHOOSING UNBIASED SAMPLES Although there are many ways of sampling a population, a random sample is preferred because it is most likely to be representative of the population.

EXAMPLE 3   Choose an unbiased sample

SENIOR CLASS PROM You are a member of the prom committee. You want to poll members of the senior class to find out where they want to hold the prom. There are 324 students in the senior class. Describe a method for selecting a random sample of 40 seniors to poll.

Solution

STEP 1   **Make** a list of all 324 seniors. Assign each senior a different integer from 1 to 324.

STEP 2   **Generate** 40 unique random integers from 1 to 324 using the *randInt* feature of a graphing calculator. The screen at the right shows six such random integers.

```
randInt(1,324)
                184
                  1
                106
                215
                 67
                213
```

If while generating the integers you obtain a duplicate, discard it and generate a new, unique integer as a replacement.

STEP 3   **Choose** the 40 students that correspond to the 40 integers you generated in Step 2.
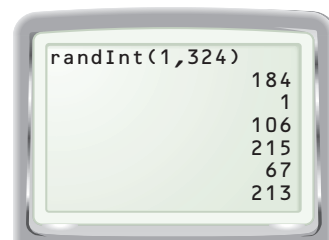
✓   **GUIDED PRACTICE**   for Examples 1, 2, and 3

1. **SCHOOL WEBSITE** A computer science teacher wants to know if students would like the morning announcements posted on the school's website. He surveys students in one of his computer science classes. Identify the type of sample described, and tell whether the sample is biased.

2. **WHAT IF?** In Example 3, what is another method you could use to generate a random sample of 40 students?

SAMPLE SIZE When conducting a survey, you need to make the size of your sample large enough so that it accurately represents the population. As the sample size increases, the *margin of error* decreases.

The **margin of error** gives a limit on how much the responses of the sample would differ from the responses of the population. For example, if 40% of the people in a poll prefer candidate A, and the margin of error is ±4%, then it is likely that between 36% and 44% of the entire population prefer candidate A.

KEY CONCEPT   *For Your Notebook*

Margin of Error Formula

When a random sample of size n is taken from a large population, the margin of error is approximated by this formula:
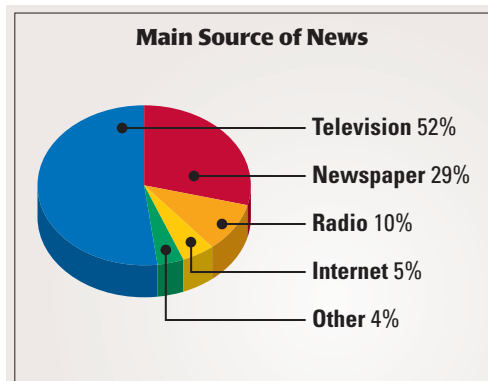
$$\text{Margin of error} = \pm\frac{1}{\sqrt{n}}$$

This means that if the percent of the sample responding a certain way is p (expressed as a decimal), then the percent of the population that would respond the same way is likely to be between $p - \dfrac{1}{\sqrt{n}}$ and $p + \dfrac{1}{\sqrt{n}}$.

EXAMPLE 4   Find a margin of error

MEDIA SURVEY In a survey of 1011 people, 52% said that television is their main source of news.

a. What is the margin of error for the survey?

b. Give an interval that is likely to contain the exact percent of all people who use television as their main source of news.

Main Source of News

- **Television** 52%
- **Newspaper** 29%
- **Radio** 10%
- **Internet** 5%
- **Other** 4%

Solution

a. Use the margin of error formula.

$$\text{Margin of error} = \pm\frac{1}{\sqrt{n}} \qquad \textcolor{blue}{\textbf{Write margin of error formula.}}$$

$$= \pm\frac{1}{\sqrt{1011}} \qquad \textcolor{blue}{\textbf{Substitute 1011 for } n.}$$

$$\approx \pm0.031 \qquad \textcolor{blue}{\textbf{Use a calculator.}}$$

▸ The margin of error for the survey is about ±3.1%.

b. To find the interval, subtract and add 3.1% to the percent of people surveyed who said television is their main source of news (52%).

$$52\% - 3.1\% = 48.9\% \qquad\qquad 52\% + 3.1\% = 55.1\%$$

▸ It is likely that the exact percent of all people who use television as their main source of news is between 48.9% and 55.1%.

EXAMPLE 5 **Standardized Test Practice**

> A polling company conducts a poll for a U.S. presidential election. How many people did the company survey if the margin of error is ±5%?
>
> **A** 25 people **B** 250 people **C** 400 people **D** 625 people

Solution

Use the margin of error formula.

$$\text{Margin of error} = \pm\frac{1}{\sqrt{n}} \qquad \textbf{Write margin of error formula.}$$

$$\pm 0.05 = \pm\frac{1}{\sqrt{n}} \qquad \textbf{Substitute ±0.05 for margin of error.}$$

$$0.0025 = \frac{1}{n} \qquad \textbf{Square each side.}$$

$$n = 400 \qquad \textbf{Solve for } \textit{n.}$$

There were 400 people surveyed.

▶ The correct answer is C. **A** **B** **C** **D**

✓ **GUIDED PRACTICE** for Examples 4 and 5

3. **INTERNET** In a survey of 1202 people, 11% said that they use the Internet or e-mail more than 10 hours per week. What is the margin of error for the survey? How many people would need to be surveyed to reduce the margin of error to ±2%?

6.4 EXERCISES

SKILL PRACTICE

1. **VOCABULARY** Copy and complete: A sample for which each member of a population has an equal chance of being selected is a(n) __?__ sample.

2. ★ **WRITING** *Describe* the difference between an unbiased sample and a biased sample.

EXAMPLES 1 and 2
for Exs. 3–5

CLASSIFYING SAMPLES **Identify the type of sample described. Then tell if the sample is biased.** *Explain* **your reasoning.**

3. A taxicab company wants to know if its customers are satisfied with the service. Each driver surveys every tenth customer during the day.

4. A town council wants to know if residents support having an off-leash area for dogs in the town park. Eighty dog owners are surveyed at the park.

5. An English teacher needs to pick 5 students to present book reports to the class. The teacher writes the names of all students in the class on pieces of paper, puts the pieces in a hat, and chooses 5 names without looking.

EXAMPLE 4
for Exs. 6–14

FINDING MARGIN OF ERROR Find the margin of error for a survey that has the given sample size. Round your answer to the nearest tenth of a percent.

6. 260   **7.** 1000   **8.** 750   **9.** 6400

10. 3275   **11.** 525   **12.** 2024   **13.** 10,000

14. ★ **MULTIPLE CHOICE** In a survey of 2000 voters, 45% said they planned to vote for candidate A. What is the margin of error for the survey?

     A ±1.8%   **B** ±2.2%   **C** ±3.6%   **D** ±4.5%

EXAMPLE 5
for Exs. 15–23

FINDING SAMPLE SIZES Find the sample size required to achieve the given margin of error. Round your answer to the nearest whole number.

15. ±3%   **16.** ±8%   **17.** ±10%   **18.** ±4.2%

19. ±5.6%   **20.** ±1.5%   **21.** ±6.5%   **22.** ±2.5%

23. ★ **MULTIPLE CHOICE** The margin of error for a poll is ±2%. What is the size of the sample?

     A 200   **B** 400   **C** 1000   **D** 2500

24. **ERROR ANALYSIS** In a survey of high school students, 13% said that they play basketball regularly. The margin of error is ±4%. *Describe* and correct the error in calculating the sample size.

$$\pm 0.13 = \pm \frac{1}{\sqrt{n}}$$
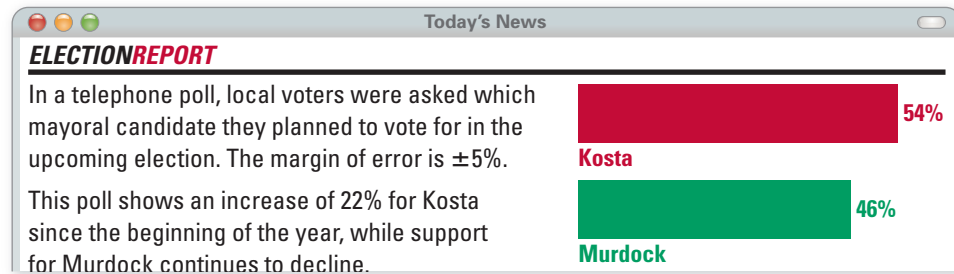$$0.0169 = \frac{1}{n}$$
$$n \approx 59$$

25. **REASONING** A survey claims the percent of a city's residents that favor building a new football stadium is likely between 52.3% and 61.7%. How many people were surveyed?

26. **CHALLENGE** Suppose a random sample of size n is required to produce a margin of error of $\pm E$. Write an expression in terms of n for the sample size needed to reduce the margin of error to $\pm \frac{1}{2}E$. By how many times must the sample size be increased in order to cut the margin of error in half?

PROBLEM SOLVING

27. **VACATION SURVEY** In a survey of 439 teenagers in the United States, 14% said that they worked during their summer vacation.

   a. What is the margin of error for the survey?

   b. Give an interval that is likely to contain the exact percent of all U.S. teenagers who worked during their summer vacation.

28. **NEWSLETTER** The staff for a student newsletter wants to conduct a survey of students' favorite TV shows. There are 1225 students in the school. The newsletter staff would like to survey 250 students. *Describe* a method for selecting an unbiased, random sample of students.

○ = See **WORKED-OUT SOLUTIONS** in Student Resources   ★ = **STANDARDIZED TEST PRACTICE**

29. ★ **SHORT RESPONSE** Based on the newspaper report shown below, is it reasonable to assume that Kosta is certain to win the election? *Explain*.

Today's News

ELECTIONREPORT

In a telephone poll, local voters were asked which mayoral candidate they planned to vote for in the upcoming election. The margin of error is ±5%.

This poll shows an increase of 22% for Kosta since the beginning of the year, while support for Murdock continues to decline.

Kosta **54%**

Murdock **46%**

30. MULTI-STEP PROBLEM A Gallup Youth Survey reported that 23% of students surveyed, or about 181 students, say that math is their favorite subject in school.

   a. How many students were surveyed?

   b. What is the margin of error for the survey?

   c. Give an interval that is likely to contain the exact percent of all students who would say that math is their favorite subject.

31. ★ **EXTENDED RESPONSE** A survey reported that 235 out of 500 voters in a sample voted for candidate A and the remainder voted for candidate B.

   a. Find Percents What percent of the voters in the sample voted for candidate A? for candidate B?

   b. Find Margin of Error What is the margin of error for the survey?

   c. Find Intervals For each candidate, find an interval that is likely to contain the exact percent of all voters who voted for the candidate.

   d. Reasoning Based on your intervals, can you be confident that candidate B won? If not, how many people in the sample would need to vote for candidate B for you to be confident of her victory? (*Hint:* Find the least number of voters for candidate B such that the intervals do not overlap.)

32. CHALLENGE In a survey, 52% of the respondents said they prefer cola X and 48% said they prefer cola Y. How many people would have to be surveyed for you to be confident that cola X is truly preferred by more than half the population? *Explain* your reasoning.

Estimate a Population Proportion

MATERIALS · graphing calculator

MATHEMATICAL PRACTICES

Reason abstractly and quantitatively.

QUESTION How can you use a sample proportion to estimate a population proportion?

You can use statistics to make reasonable predictions, or *inferences*, about an entire population from a sample of the population. A *population proportion* is the ratio of members of a population with a particular characteristic to total members of the population. A *sample proportion* is the ratio of members of a sample of the population with a particular characteristic to total members of the sample.

EXPLORE 1 Compare sample proportions from a population with known proportion

In the "population" of randomly-generated decimals between 0 and 1, the proportion of decimals less than 0.6 is 0.6. Simulate random samples of decimals from this population, and compare the sample proportions with the population proportion.

STEP 1 *Perform a simulation*

Use a graphing calculator to generate 40 random decimals between 0 and 1. Press [MATH], choose the "PRB" menu, select "rand," and press [ENTER]. The display will be similar to the one at the right.

```
rand
         .8999652259
```
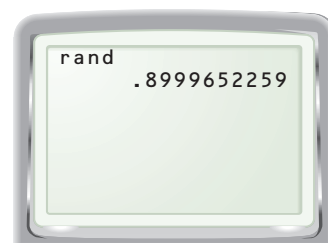
Press [ENTER] 40 times to generate 40 random numbers. Keep a tally of the number of decimals below 0.6. Then find your sample proportion by dividing the number of decimals below 0.6—the "successes"—by the number of trials, which is 40.

Repeat the simulation to find another sample proportion. Each group will need to do enough simulations so that there are 40 in all for the class.

STEP 2 *Display sample proportions*

Collect all 40 sample proportions. Make a histogram of the sample proportions using intervals of 0.05, for example, [0.55, 0.6). The histogram for one set of 40 sample proportions is shown at the right.
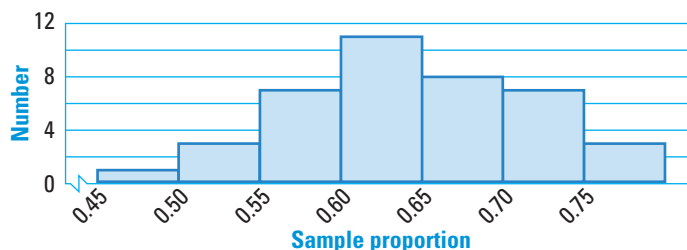
STEP 3 *Investigate sample proportions*

Exclude the single smallest and single greatest of the 40 sample proportions. Then find the interval that contains the remaining 38 sample proportions.

Plot the interval containing the middle 38 sample proportions as shown. The blue horizontal segment represents the middle 38 of 40, or 95%, of the sample proportions. So, 95% of the time, the sample proportion from the simulation fell within the interval graphed.
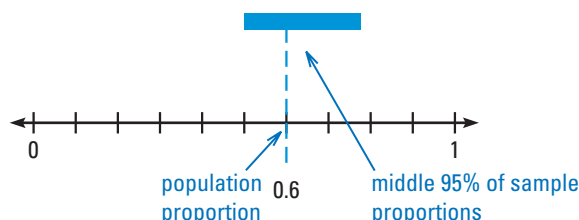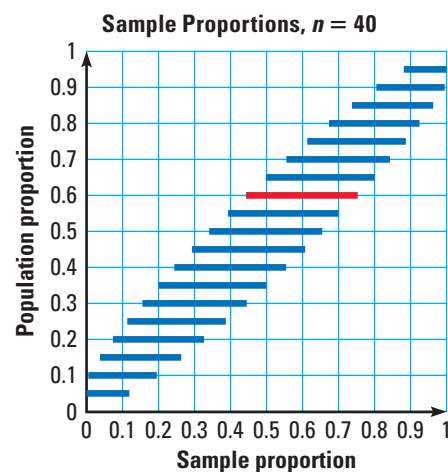
EXPLORE 2 · Estimate a population proportion from a sample proportion

Sample Proportions, n = 40

STEP 1 · Investigate sample proportions

There are formulas for calculating intervals expected to contain 95% of the sample proportions given any population proportion and sample size. The chart at the right shows the 95% sample proportion intervals for population proportions of 0.05, 0.1, 0.15, ..., 0.9, 0.95 where the sample size is $n = 40$.

The interval in red is the calculated 95% sample proportion interval for a population proportion of 0.6. Compare this interval with the 95% interval your class created for the simulation in Explore 1.

STEP 2 · Use a sample proportion

It's not usually possible to collect data for entire populations. So, data from samples are used to estimate unknown population proportions. Suppose you want to estimate a population proportion using a sample of size 40, and your sample has a sample proportion of 0.6. Which populations are most likely to have produced a sample like yours?

The green vertical line shown at right highlights a sample proportion of 0.6. Any horizontal bar that this line touches corresponds to a population that will produce a sample whose sample proportion lies within that bar 95% of the time. The green line touches the bars for population proportions of 0.45 up to 0.75. So, *if we have a sample of size 40 with a sample proportion of 0.6, then, 95% of the time, the population proportion should lie in the interval from 0.45 to 0.75.*
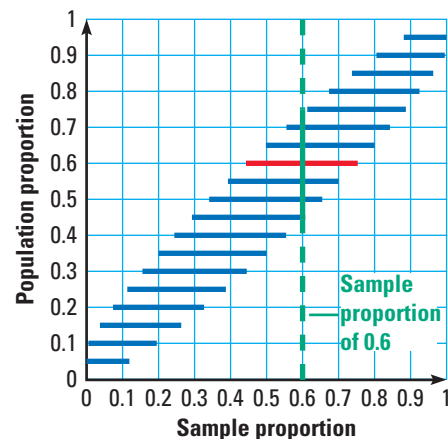
STEP 3 · Estimate a population proportion

For a sample of size 40 with sample proportion 0.3, use the chart to estimate the interval that contains the corresponding population proportion 95% of the time.

DRAW CONCLUSIONS · Use your observations to complete these exercises

1. As a class, repeat Steps 1 and 2 of Explore 1, but with each trial count the number of decimals less than 0.6 out of *80* random decimals. *Describe* how increasing the sample size in each trial affects the shape of the histogram.

2. You have used the *margin of error* $\pm \frac{1}{\sqrt{n}}$ for a sample of size n to estimate an interval containing a population proportion. More precisely, this expression is $\pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where the symbol $\hat{p}$ (read as "p-hat") is the sample proportion. Then $\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is a 95% *confidence interval* for the population proportion p. We can expect the population proportion to be in this interval 95% of the time. Find a 95% confidence interval for the population proportion for a sample of size 150 with a sample proportion of 0.8.