- Other things being equal, the **margin of error** of a confidence interval gets smaller as
  - the confidence level C decreases.
  - the sample size n increases.
- Remember that the margin of error for a confidence interval includes only chance variation, not other sources of error like nonresponse and undercoverage.

Section 8.1 Exercises

In Exercises 1 to 4, determine the point estimator you would use and calculate the value of the point estimate.

1. **Got shoes?** How many pairs of shoes, on average, do female teens have? To find out, an AP® Statistics class conducted a survey. They selected an SRS of 20 female students from their school. Then they recorded the number of pairs of shoes that each student reported having. Here are the data:

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

2. **Got shoes?** The class in Exercise 1 wants to estimate the variability in the number of pairs of shoes that female students have by estimating the population variance σ^2.

3. **Going to the prom** Tonya wants to estimate what proportion of the seniors in her school plan to attend the prom. She interviews an SRS of 50 of the 750 seniors in her school and finds that 36 plan to go to the prom.

4. **Reporting cheating** What proportion of students are willing to report cheating by other students? A student project put this question to an SRS of 172 undergraduates at a large university: "You witness two students cheating on a quiz. Do you go to the professor?" Only 19 answered "Yes."[3]

5. **NAEP scores** Young people have a better chance of full-time employment and good wages if they are good with numbers. How strong are the quantitative skills of young Americans of working age? One source of data is the National Assessment of Educational Progress (NAEP) Young Adult Literacy Assessment Survey, which is based on a nationwide probability sample of households. The NAEP survey includes a short test of quantitative skills, covering mainly basic arithmetic and the ability to apply it to realistic problems. Scores on the test range from 0 to 500. For example, a person who scores 233 can add the amounts of two checks appearing on a bank deposit slip; someone scoring 325 can determine the price of a meal from a menu; a person scoring 375 can transform a price in cents per ounce into dollars per pound.[4]

   Suppose that you give the NAEP test to an SRS of 840 people from a large population in which the scores have mean 280 and standard deviation $\sigma = 60$. The mean $\bar{x}$ of the 840 scores will vary if you take repeated samples.

   (a) Describe the shape, center, and spread of the sampling distribution of $\bar{x}$.

   (b) Sketch the sampling distribution of $\bar{x}$. Mark its mean and the values 1, 2, and 3 standard deviations on either side of the mean.

   (c) According to the 68–95–99.7 rule, about 95% of all values of $\bar{x}$ lie within a distance m of the mean of the sampling distribution. What is m? Shade the region *on the axis* of your sketch that is within m of the mean.

   (d) Whenever $\bar{x}$ falls in the region you shaded, the population mean μ lies in the confidence interval $\bar{x} \pm m$. For what percent of all possible samples does the interval capture μ?

6. **Auto emissions** Oxides of nitrogen (called NOX for short) emitted by cars and trucks are important contributors to air pollution. The amount of NOX emitted by a particular model varies from vehicle to vehicle. For one light-truck model, NOX emissions vary with mean $\mu = 1.8$ grams per mile and standard deviation $\sigma = 0.4$ gram per mile. You test an SRS of 50 of these trucks. The sample mean NOX level $\bar{x}$ will vary if you take repeated samples.

(a) Describe the shape, center, and spread of the sampling distribution of $\bar{x}$.

(b) Sketch the sampling distribution of $\bar{x}$. Mark its mean and the values 1, 2, and 3 standard deviations on either side of the mean.

(c) According to the 68–95–99.7 rule, about 95% of all values of $\bar{x}$ lie within a distance m of the mean of the sampling distribution. What is m? Shade the region *on the axis* of your sketch that is within m of the mean.

(d) Whenever $\bar{x}$ falls in the region you shaded, the unknown population mean μ lies in the confidence interval $\bar{x} \pm m$. For what percent of all possible samples does the interval capture μ?

7. **NAEP scores** Refer to Exercise 5. Below your sketch, choose one value of $\bar{x}$ inside the shaded region and draw its corresponding confidence interval. Do the same for one value of $\bar{x}$ outside the shaded region. What is the most important difference between these intervals? (Use Figure 8.5, on page 483, as a model for your drawing.)

8. **Auto emissions** Refer to Exercise 6. Below your sketch, choose one value of $\bar{x}$ inside the shaded region and draw its corresponding confidence interval. Do the same for one value of $\bar{x}$ outside the shaded region. What is the most important difference between these intervals? (Use Figure 8.5, on page 483, as a model for your drawing.)

9. pg 481 **Prayer in school** A *New York Times*/CBS News Poll asked a random sample of U.S. adults the question, "Do you favor an amendment to the Constitution that would permit organized prayer in public schools?" Based on this poll, the 95% confidence interval for the population proportion who favor such an amendment is (0.63, 0.69).

(a) Interpret the confidence interval.

(b) What is the point estimate that was used to create the interval? What is the margin of error?

(c) Based on this poll, a reporter claims that more than two-thirds of U.S. adults favor such an amendment. Use the confidence interval to evaluate this claim.

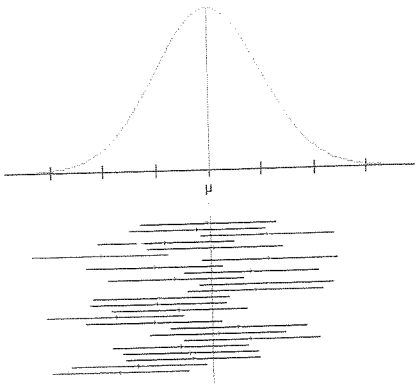
10. **Losing weight** A Gallup Poll asked a random sample of U.S. adults, "Would you like to lose weight?" Based on this poll, the 95% confidence interval for the population proportion who want to lose weight is (0.56, 0.62).[5]

(a) Interpret the confidence interval.

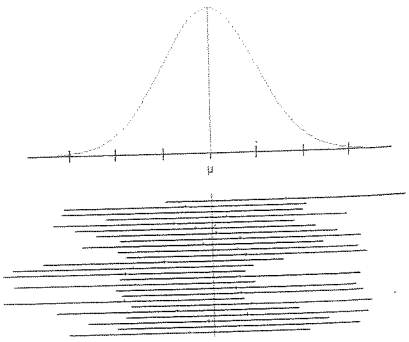
(b) What is the point estimate that was used to create the interval? What is the margin of error?

(c) Based on this poll, Gallup claims that more than half of U.S. adults want to lose weight. Use the confidence interval to evaluate this claim.

11. **How confident?** The figure below shows the result of taking 25 SRSs from a Normal population and constructing a confidence interval for each sample. Which confidence level—80%, 90%, 95%, or 99%—do you think was used? Explain.



12. **How confident?** The figure below shows the result of taking 25 SRSs from a Normal population and constructing a confidence interval for each sample. Which confidence level—80%, 90%, 95%, or 99%—do you think was used? Explain.



13. **Prayer in school** Refer to Exercise 9. The news article goes on to say: "The theoretical errors do not take into account · · · additional error resulting from the various practical difficulties in taking any survey of public opinion." List some of the "practical difficulties" that may cause errors which are not included in the ±3 percentage point margin of error.

14. **Losing weight** Refer to Exercise 10. As Gallup indicates, the 3 percentage point margin of error for this poll includes only sampling variability (what they call "sampling error"). What other potential sources of error (Gallup calls these "nonsampling errors") could affect the accuracy of the 95% confidence interval?

15. pg 484 **Shoes** The AP® Statistics class in Exercise 1 also asked an SRS of 20 boys at their school how many

pairs of shoes they have. A 95% confidence interval for the difference in the population means (girls – boys) is 10.9 to 26.5. Interpret the confidence interval and the confidence level.

Lying online Many teens have posted profiles on sites such as Facebook. A sample survey asked random samples of teens with online profiles if they included false information in their profiles. Of 170 younger teens (ages 12 to 14) polled, 117 said "Yes." Of 317 older teens (ages 15 to 17) polled, 152 said "Yes." A 95% confidence interval for the difference in the population proportions (younger teens – older teens) is 0.120 to 0.297. Interpret the confidence interval and the confidence level.

Shoes Refer to Exercise 15. Does the confidence interval give convincing evidence of a difference in the population mean number of pairs of shoes for boys and girls at the school? Justify your answer.

Lying online Refer to Exercise 16. Does the confidence interval give convincing evidence of a difference in the population proportions of younger and older teens who include false information in their profiles? Justify your answer.

Explaining confidence A 95% confidence interval for the mean body mass index (BMI) of young American women is 26.8 ± 0.6. Discuss whether each of the following explanations is correct.

We are confident that 95% of all young women have BMI between 26.2 and 27.4.

We are 95% confident that future samples of young women will have mean BMI between 26.2 and 27.4.

Any value from 26.2 to 27.4 is believable as the true mean BMI of young American women.

If we take many samples, the population mean BMI will be between 26.2 and 27.4 in about 95% of those samples.

The mean BMI of young American women cannot be 28.

Explaining confidence The admissions director from Big City University found that (107.8, 116.2) is a 95% confidence interval for the mean IQ score of all freshmen. Discuss whether each of the following explanations is correct.

There is a 95% probability that the interval from 107.8 to 116.2 contains μ.

There is a 95% chance that the interval (107.8, 116.2) contains $\bar{x}$.

This interval was constructed using a method that produces intervals that capture the true mean in 95% of all possible samples.

(d) If we take many samples, about 95% of them will contain the interval (107.8, 116.2).

(e) The probability that the interval (107.8, 116.2) captures μ is either 0 or 1, but we don't know which.

Multiple choice: Select the best answer for Exercises 21 to 24.

Exercises 21 and 22 refer to the following setting. A researcher plans to use a random sample of families to estimate the mean monthly family income for a large population.

21. The researcher is deciding between a 95% confidence level and a 99% confidence level. Compared to a 95% confidence interval, a 99% confidence interval will be

(a) narrower and would involve a larger risk of being incorrect.

(b) wider and would involve a smaller risk of being incorrect.

(c) narrower and would involve a smaller risk of being incorrect.

(d) wider and would involve a larger risk of being incorrect.

(e) wider and would have the same risk of being incorrect.

22. The researcher is deciding between a sample of size $n = 500$ and a sample of size $n = 1000$. Compared to using a sample size of $n = 500$, a 95% confidence interval based on a sample size of $n = 1000$ will be

(a) narrower and would involve a larger risk of being incorrect.

(b) wider and would involve a smaller risk of being incorrect.

(c) narrower and would involve a smaller risk of being incorrect.

(d) wider and would involve a larger risk of being incorrect.

(e) narrower and would have the same risk of being incorrect.

23. In a poll,

I. Some people refused to answer questions.

II. People without telephones could not be in the sample.

III. Some people never answered the phone in several calls.

Which of these possible sources of bias is included in the $\pm 2\%$ margin of error announced for the poll?

(a) I only (c) III only (e) None of these

(b) II only (d) I, II, and III

24. You have measured the systolic blood pressure of an SRS of 25 company employees. A 95% confidence interval for the mean systolic blood pressure for the employees of this company is (122, 138). Which of the following statements is true?

(a) 95% of the sample of employees have a systolic blood pressure between 122 and 138.

(b) 95% of the population of employees have a systolic blood pressure between 122 and 138.

(c) If the procedure were repeated many times, 95% of the resulting confidence intervals would contain the population mean systolic blood pressure.

(d) If the procedure were repeated many times, 95% of the time the population mean systolic blood pressure would be between 122 and 138.

(e) If the procedure were repeated many times, 95% of the time the sample mean systolic blood pressure would be between 122 and 138.

25. **Power lines and cancer** (4.2, 4.3) Does living near power lines cause leukemia in children? The National Cancer Institute spent 5 years and $5 million gathering data on this question. The researchers compared 638 children who had leukemia with 620 who did not. They went into the homes and measured the magnetic fields in children's bedrooms, in other rooms, and at the front door. They recorded facts about power lines near the family home and also near the mother's residence when she was pregnant. Result: no connection between leukemia and exposure to magnetic fields of the kind produced by power lines was found.[7]

(a) Was this an observational study or an experiment? Justify your answer.

(b) Does this study show that living near power lines doesn't cause cancer? Explain.

26. **Sisters and brothers** (3.1, 3.2) How strongly do physical characteristics of sisters and brothers correlate? Here are data on the heights (in inches) of 11 adult pairs:[8]

Brother:	71	68	66	67	70	71	70	73	72	65	66
Sister:	69	64	65	63	65	62	65	64	66	59	62

(a) Construct a scatterplot using brother's height as the explanatory variable. Describe what you see.

(b) Use your calculator to compute the least-squares regression line for predicting sister's height from brother's height. Interpret the slope in context.

(c) Damien is 70 inches tall. Predict the height of his sister Tonya.

(d) Do you expect your prediction in (c) to be very accurate? Give appropriate evidence to support your answer.

8.2 Estimating a Population Proportion

WHAT YOU WILL LEARN By the end of the section, you should be able to:

- State and check the Random, 10%, and Large Counts conditions for constructing a confidence interval for a population proportion.
- Determine critical values for calculating a C% confidence interval for a population proportion using a table or technology.
- Construct and interpret a confidence interval for a population proportion.
- Determine the sample size required to obtain a C% confidence interval for a population proportion with a specified margin of error.

In Section 8.1, we saw that a confidence interval can be used to estimate an unknown population parameter. We are often interested in estimating the proportion p of some outcome in the population. Here are some examples:

- What proportion of U.S. adults are unemployed right now?
- What proportion of high school students have cheated on a test?