Section 7.1 Summary

- A **parameter** is a number that describes a population. To estimate an unknown parameter, use a **statistic** calculated from a sample.
- The **population distribution** of a variable describes the values of the variable for all individuals in a population. The **sampling distribution** of a statistic describes the values of the statistic in all possible samples of the same size from the same population. Don't confuse the sampling distribution with a **distribution of sample data**, which gives the values of the variable for all individuals in a particular sample.
- A statistic can be an **unbiased estimator** or a **biased estimator** of a parameter. A statistic is a biased estimator if the center (mean) of its sampling distribution is not equal to the true value of the parameter.
- The **variability** of a statistic is described by the spread of its sampling distribution. Larger samples give smaller spread.
- When trying to estimate a parameter, choose a statistic with low or no bias and minimum variability.

Section 7.1 Exercises

For Exercises 1 and 2, identify the population, the parameter, the sample, and the statistic in each setting.

1. Healthy living

pg 425

(a) A random sample of 1000 people who signed a card saying they intended to quit smoking were contacted 9 months later. It turned out that 210 (21%) of the sampled individuals had not smoked over the past 6 months.

(b) Tom is cooking a large turkey breast for a holiday meal. He wants to be sure that the turkey is safe to eat, which requires a minimum internal temperature of 165°F. Tom uses a thermometer to measure the temperature of the turkey meat at four randomly chosen points. The minimum reading in the sample is 170°F.

2. The economy

(a) Each month, the Current Population Survey interviews a random sample of individuals in about 60,000 U.S. households. One of their goals is to estimate the national unemployment rate. In October 2012, 7.9% of those interviewed were unemployed.

(b) How much do gasoline prices vary in a large city? To find out, a reporter records the price per gallon of regular unleaded gasoline at a random sample of 10 gas stations in the city on the same day. The range (maximum − minimum) of the prices in the sample is 25 cents.

For each boldface number in Exercises 3 to 6, (1) state whether it is a parameter or a statistic and (2) use appropriate notation to describe each number; for example, $p = 0.65$.

3. **Get your bearings** A large container is full of ball bearings with mean diameter 2.5003 centimeters (cm). This is within the specifications for acceptance of the container by the purchaser. By chance, an inspector chooses 100 bearings from the container that have mean diameter 2.5009 cm. Because this is outside the specified limits, the container is mistakenly rejected.

4. **Voters** Voter registration records show that 41% of voters in a state are registered as Democrats. To test a random digit dialing device, you use it to call 250 randomly chosen residential telephones in the state. Of the registered voters contacted, 33% are registered Democrats.

5. **Unlisted numbers** A telemarketing firm in a large city uses a device that dials residential telephone numbers in that city at random. Of the first 100 numbers dialed, 48% are unlisted. This is not surprising because 52% of all residential phones in the city are unlisted.

6. **How tall?** A random sample of female college students has a mean height of 64.5 inches, which is greater than the 63-inch mean height of all adult American women.

Exercises 7 and 8 refer to the small population {2, 6, 8, 10, 10, 12} with mean $\mu = 8$ and range 10.

7. **Sampling distribution**

(a) List all 15 possible SRSs of size $n = 2$ from the population. Find the value of $\bar{x}$ for each sample.

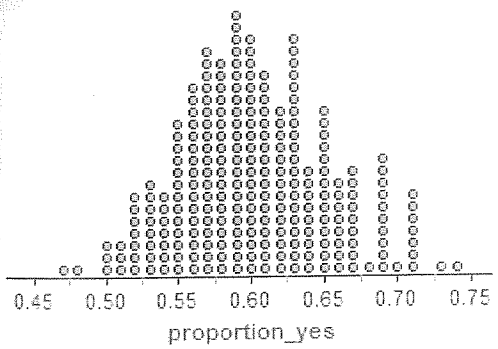
(b) Make a graph of the sampling distribution of $\bar{x}$. Describe what you see.

8. **Sampling distribution**

(a) List all 15 possible SRSs of size $n = 2$ from the population. Find the value of the range for each sample.

(b) Make a graph of the sampling distribution of the sample range. Describe what you see.

9. **Doing homework** A school newspaper article claims that 60% of the students at a large high school did all their assigned homework last week. Some skeptical AP® Statistics students want to investigate whether this claim is true, so they choose an SRS of 100 students from the school to interview. What values of the sample proportion $\hat{p}$ would be consistent with the claim that the population proportion of students who completed all their homework is $p = 0.60$? To find out, we used Fathom software to simulate choosing 250 SRSs of size $n = 100$ students from a population in which $p = 0.60$. The figure below is a dotplot of the sample proportion $\hat{p}$ of students who did all their homework.



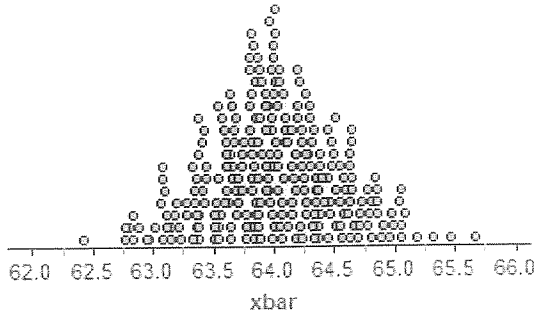
(a) There is one dot on the graph at 0.73. Explain what this value represents.

(b) Describe the distribution. Are there any obvious outliers?

(c) Would it be surprising to get a sample proportion of 0.45 or lower in an SRS of size 100 when $p = 0.6$? Justify your answer.

(d) Suppose that 45 of the 100 students in the actual sample say that they did all their homework last week. What would you conclude about the newspaper article's claim? Explain.

10. **Tall girls** According to the National Center for Health Statistics, the distribution of heights for 16-year-old females is modeled well by a Normal density curve with mean $\mu = 64$ inches and standard deviation $\sigma = 2.5$ inches. To see if this distribution applies at their high school, an AP® Statistics class takes an SRS of 20 of the 300 16-year-old females at the school and measures their heights. What values of the sample mean $\bar{x}$ would be consistent with the population distribution being $N(64, 2.5)$? To find out, we used Fathom software to simulate choosing 250 SRSs of size $n = 20$ students from a population that is $N(64, 2.5)$. The figure below is a dotplot of the sample mean height $\bar{x}$ of the students in each sample.



(a) There is one dot on the graph at 62.4. Explain what this value represents.

(b) Describe the distribution. Are there any obvious outliers?

(c) Would it be surprising to get a sample mean of 64.7 or more in an SRS of size 20 when $\mu = 64$? Justify your answer.

(d) Suppose that the average height of the 20 girls in the class's actual sample is $\bar{x} = 64.7$. What would you conclude about the population mean height μ for the 16-year-old females at the school? Explain.

11. **Doing homework** Refer to Exercise 9.

(a) Make a bar graph of the population distribution given that the newspaper's claim is correct.

(b) Sketch a possible graph of the distribution of sample data for the SRS of size 100 taken by the AP® Statistics students.

12. **Tall girls** Refer to Exercise 10.

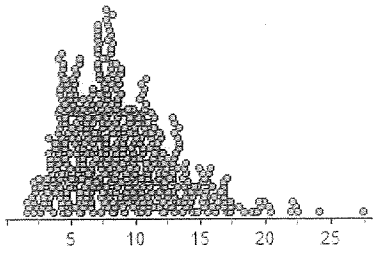
(a) Make a graph of the population distribution.

(b) Sketch a possible dotplot of the distribution of sample data for the SRS of size 20 taken by the AP® Statistics class.

Exercises 13 and 14 refer to the following setting. During the winter months, outside temperatures at the Starneses' cabin in Colorado can stay well below freezing (32°F, or 0°C) for weeks at a time. To prevent the pipes from freezing, Mrs. Starnes sets the thermostat at 50°F. The manufacturer claims that the thermostat allows variation in home temperature that follows a Normal distribution with $\sigma = 3$°F. To test this claim, Mrs. Starnes programs her digital thermometer to take an SRS of $n = 10$ readings during a 24-hour period. Suppose the thermostat is

working properly and that the actual temperatures in the cabin vary according to a Normal distribution with mean $\mu = 50°F$ and standard deviation $\sigma = 3°F$.

13. **Cold cabin?** The Fathom screen shot below shows the results of taking 500 SRSs of 10 temperature readings from a population distribution that is $N(50, 3)$ and recording the sample variance s_x^2 each time.

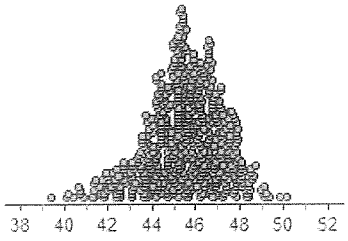


Sample variance

(a) Describe the approximate sampling distribution.

(b) Suppose that the variance from an actual sample is $s_x^2 = 25$. What would you conduct about the thermostat manufacturer's claim? Explain.

14. **Cold cabin?** The Fathom screen shot below shows the results of taking 500 SRSs of 10 temperature readings from a population distribution that is $N(50, 3)$ and recording the sample minimum each time.



Sample minimum

(a) Describe the approximate sampling distribution.

(b) Suppose that the minimum of an actual sample is 40°F. What would you conclude about the thermostat manufacturer's claim? Explain.

15. **A sample of teens** A study of the health of teenagers plans to measure the blood cholesterol levels of an SRS of 13- to 16-year-olds. The researchers will report the mean $\bar{x}$ from their sample as an estimate of the mean cholesterol level μ in this population. Explain to someone who knows little about statistics what it means to say that $\bar{x}$ is an unbiased estimator of μ.

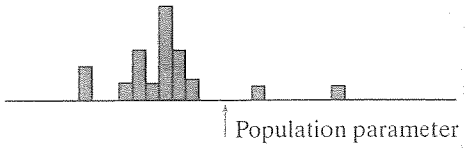
16. **Predict the election** A polling organization plans to ask a random sample of likely voters who they plan to vote for in an upcoming election. The researchers will report the sample proportion $\hat{p}$ that favors the incumbent as an estimate of the population proportion p that favors the incumbent. Explain to someone who knows little about statistics what it means to say that $\hat{p}$ is an unbiased estimator of p.

17. **A sample of teens** Refer to Exercise 15. The sample mean $\bar{x}$ is an unbiased estimator of the population mean μ no matter what size SRS the study chooses. Explain to someone who knows nothing about statistics why a large random sample will give more trustworthy results than a small random sample.

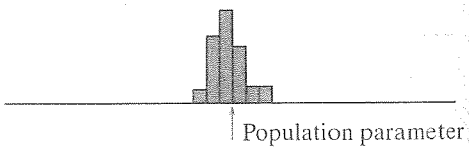
18. **Predict the election** Refer to Exercise 16. The sample proportion $\hat{p}$ is an unbiased estimator of the population proportion p no matter what size random sample the polling organization chooses. Explain to someone who knows nothing about statistics why a large random sample will give more trustworthy results than a small random sample.

19. **Bias and variability** The figure below shows histograms of four sampling distributions of different statistics intended to estimate the same parameter.

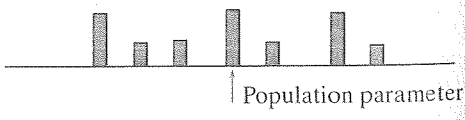
pg 435



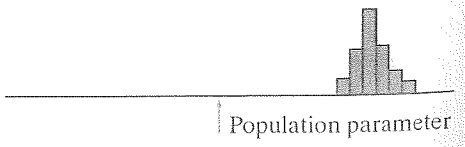
(i)



(ii)



(iii)



(iv)

(a) Which statistics are unbiased estimators? Justify your answer.

(b) Which statistic does the best job of estimating the parameter? Explain.

20. **IRS audits** The Internal Revenue Service plans to examine an SRS of individual federal income tax returns. The parameter of interest is the proportion of all returns claiming itemized deductions. Which would be better for estimating this parameter: an SRS of 20,000 returns or an SRS of 2000 returns? Justify your answer.

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

21. At a particular college, 78% of all students are receiving some kind of financial aid. The school newspaper selects a random sample of 100 students and 72% of the respondents say they are receiving some sort of financial aid. Which of the following is true?

(a) 78% is a population and 72% is a sample.

(b) 72% is a population and 78% is a sample.

(c) 78% is a parameter and 72% is a statistic.

(d) 72% is a parameter and 78% is a statistic.

(e) 78% is a parameter and 100 is a statistic.

22. A statistic is an unbiased estimator of a parameter when

(a) the statistic is calculated from a random sample.

(b) in a single sample, the value of the statistic is equal to the value of the parameter.

(c) in many samples, the values of the statistic are very close to the value of the parameter.

(d) in many samples, the values of the statistic are centered at the value of the parameter.

(e) in many samples, the distribution of the statistic has a shape that is approximately Normal.

23. In a residential neighborhood, the median value of a house is $200,000. For which of the following sample sizes is the sample median most likely to be above $250,000?

(a) $n = 10$

(b) $n = 50$

(c) $n = 100$

(d) $n = 1000$

(e) Impossible to determine without more information.

24. Increasing the sample size of an opinion poll will reduce the

(a) bias of the estimates made from the data collected in the poll.

(b) variability of the estimates made from the data collected in the poll.

(c) effect of nonresponse on the poll.

(d) variability of opinions in the sample.

(e) variability of opinions in the population.

25. **Dem bones** (2.2) Osteoporosis is a condition in which the bones become brittle due to loss of minerals. To diagnose osteoporosis, an elaborate apparatus measures bone mineral density (BMD). BMD is usually reported in standardized form. The standardization is based on a population of healthy young adults. The World Health Organization (WHO) criterion for osteoporosis is a BMD score that is 2.5 standard deviations below the mean for young adults. BMD measurements in a population of people similar in age and gender roughly follow a Normal distribution.

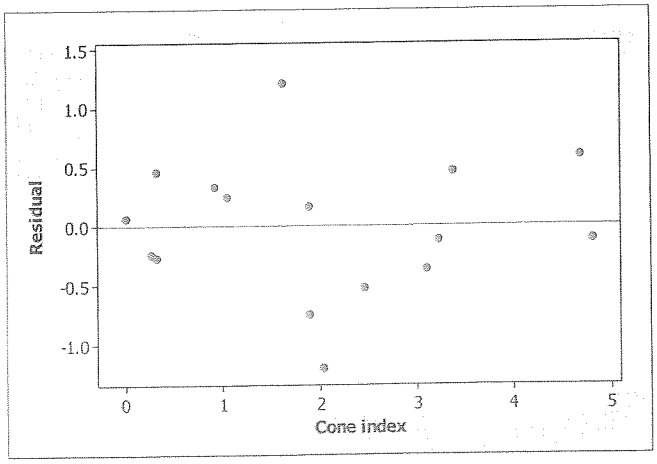
(a) What percent of healthy young adults have osteoporosis by the WHO criterion?

(b) Women aged 70 to 79 are, of course, not young adults. The mean BMD in this age group is about -2 on the standard scale for young adults. Suppose that the standard deviation is the same as for young adults. What percent of this older population has osteoporosis?

26. **Squirrels and their food supply** (3.2) Animal species produce more offspring when their supply of food goes up. Some animals appear able to anticipate unusual food abundance. Red squirrels eat seeds from pinecones, a food source that sometimes has very large crops. Researchers collected data on an index of the abundance of pinecones and the average number of offspring per female over 16 years.[3] Computer output from a least-squares regression on these data and a residual plot are shown below.

Predictor	Coef	SE Coef	T	P
Constant	1.4146	0.2517	5.62	0.000
Cone index	0.4399	0.1016	4.33	0.001

S = 0.600309 R-Sq = 57.2% R-Sq(adj) = 54.2%



(a) Give the equation for the least-squares regression line. Define any variables you use.

(b) Is a linear model appropriate for these data? Explain.

(c) Interpret the values of r^2 and s in context.