

The lesson about center and spread is clear: given a choice of statistics to estimate an unknown parameter, choose one with no or low bias and minimum variability. Shape is a more complicated issue. We have seen sampling distributions that are left-skewed, right-skewed, roughly symmetric, and even approximately Normal. The same statistic can have sampling distributions with different shapes depending on the population distribution and the sample size. Our advice: be sure to consider the shape of the sampling distribution before doing inference.

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 sample.
- A statistic can be an **unbiased estimator** or a **biased estimator** of a parameter. Bias means that the center (mean) of the 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. Don't forget to consider the shape of the sampling distribution before doing inference.

SECTION 7.1

Exercises

For Exercises 1 to 4, identify the population, the parameter, the sample, and the statistic in each setting.

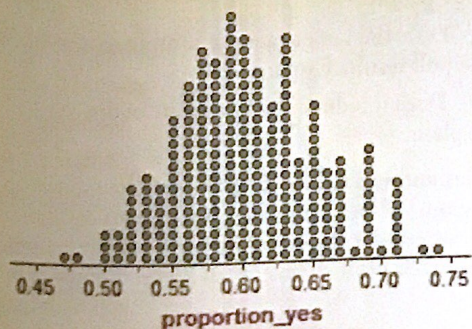
pg 416

1. **Stop smoking!** A random sample of 1000 people who signed a card saying they intended to quit smoking were contacted nine months later. It turned out that 210 (21%) of the sampled individuals had not smoked over the past six months.
2. **Unemployment** Each month, the Current Population Survey interviews a random sample of individuals in about 55,000 U.S. households. One of their goals is to estimate the national unemployment rate. In December 2009, 10.0% of those interviewed were unemployed.
3. **Hot turkey** 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.
4. **Gas prices** 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 5 to 8, (1) state whether it is a parameter or a statistic and (2) use appropriate notation to describe each number; for example, $p = 0.65$.

5. **Get your bearings** A large container of ball bearings has 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.
6. **Florida voters** Florida has played a key role in recent presidential elections. Voter registration records show that 41% of Florida voters are registered as Democrats. To test a random digit dialing device, you use it to call 250 randomly chosen residential telephones in Florida. Of the registered voters contacted, 33% are registered Democrats.
7. **Unlisted numbers** A telemarketing firm in Los Angeles 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 Los Angeles residential phones are unlisted.
8. **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.

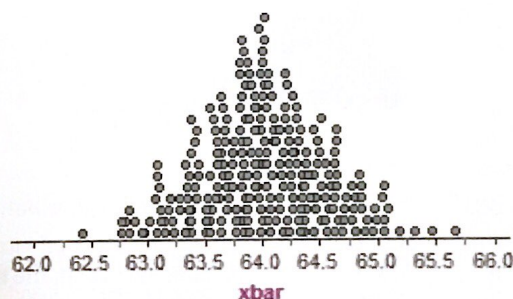
9. **Doing homework** A school newspaper article claims that 60% of the students at a large high school did all their assigned homework last week. Some skeptical AP Statistics students want to investigate whether this claim is true, so they choose an SRS of 100 students from the school to interview. 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) Is this the sampling distribution of \hat{p} ? Justify your answer.

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



- (a) Is this the sampling distribution of \bar{x} ? Justify your answer.
- (b) Describe the distribution. Are there any obvious outliers?
- (c) 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.

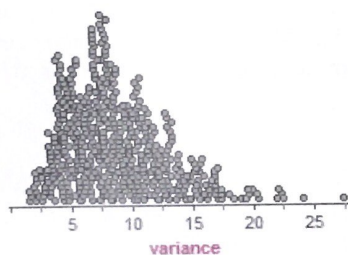
- (a) Make a graph of the population distribution given that the newspaper's claim is correct and that there are 3000 students in the school.
(Hint: What type of variable is being measured?)
- (b) Sketch a possible graph of the distribution of sample data for the SRS of size 100 taken by the AP Statistics students.

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

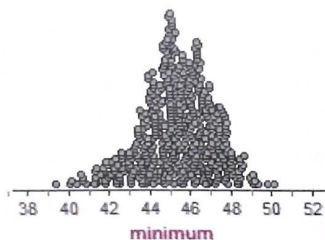
- (a) Make a graph of the population distribution.
- (b) Sketch a possible graph of the distribution of sample data for the SRS of size 20 taken by the AP Statistics class.

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 $\sigma = 3^{\circ}\text{F}$. 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 $\mu = 50^{\circ}\text{F}$ and standard deviation $\sigma = 3^{\circ}\text{F}$.

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



- (a) Describe the approximate sampling distribution.
 (b) Suppose that the variance from an actual sample is $s_x^2 = 25$. What would you conclude 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's $N(50, 3)$ and recording the sample minimum each time.



- (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. **Run a mile** During World War II, 12,000 able-bodied male undergraduates at the University of Illinois participated in required physical training. Each student ran a timed mile. Their times followed the Normal distribution with mean 7.11 minutes and

standard deviation 0.74 minute. An SRS of 100 of these students has mean time $\bar{x} = 7.15$ minutes. A second SRS of size 100 has mean $\bar{x} = 6.97$ minutes. After many SRSs, the values of the sample mean \bar{x} follow the Normal distribution with mean 7.11 minutes and standard deviation 0.074 minute.

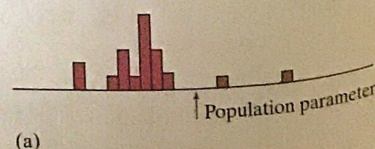
- (a) What is the population? Describe the population distribution.
 (b) Describe the sampling distribution of \bar{x} . How is it different from the population distribution?
16. **Scooping beads** A statistics teacher fills a large container with 1000 white and 3000 red beads and then mixes the beads thoroughly. She then has her students take repeated SRSs of 50 beads from the container. After many SRSs, the values of the sample proportion \hat{p} of red beads are approximated well by a Normal distribution with mean 0.75 and standard deviation 0.06.
- (a) What is the population? Describe the population distribution.
 (b) Describe the sampling distribution of \hat{p} . How is it different from the population distribution?
17. **IRS audits** The Internal Revenue Service plans to examine an SRS of individual federal income tax returns from each state. One variable of interest is the proportion of returns claiming itemized deductions. The total number of tax returns in each state varies from over 15 million in California to about 240,000 in Wyoming.
- (a) Will the sampling variability of the sample proportion change from state to state if an SRS of 2000 tax returns is selected in each state? Explain your answer.
 (b) Will the sampling variability of the sample proportion change from state to state if an SRS of 1% of all tax returns is selected in each state? Explain your answer.

18. **Predict the election** Just before a presidential election, a national opinion poll increases the size of its weekly random sample from the usual 1500 people to 4000 people.

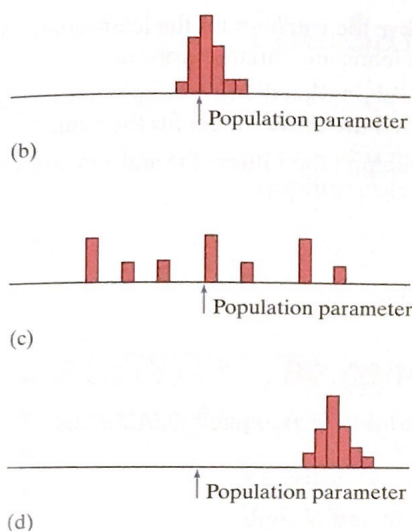
- (a) Does the larger random sample reduce the bias of the poll result? Explain.
 (b) Does it reduce the variability of the result? Explain.

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

pg 427



(a)



- (a) Which statistics are unbiased estimators? Justify your answer.
- (b) Which statistic does the best job of estimating the parameter? Explain.

20. **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.

- (a) Explain to someone who knows no statistics what it means to say that \bar{x} is an unbiased estimator of μ .
- (b) The sample result \bar{x} is an unbiased estimator of the population mean μ no matter what size SRS the study chooses. Explain to someone who knows no statistics why a large random sample gives more trustworthy results than a small random sample.

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

21. A newspaper poll reported that 73% of respondents liked business tycoon Donald Trump. The number 73% is

- (a) a population.
- (b) a parameter.
- (c) a sample.
- (d) a statistic.
- (e) an unbiased estimator.

22. The name for the pattern of values that a statistic takes when we sample repeatedly from the same population is

- (a) the bias of the statistic.
- (b) the variability of the statistic.
- (c) the population distribution.

- (d) the distribution of sample data.
- (e) the sampling distribution of the statistic.

23. If we take a simple random sample of size $n = 500$ from a population of size 5,000,000, the variability of our estimate will be

- (a) much less than the variability for a sample of size $n = 500$ from a population of size 50,000,000.
- (b) slightly less than the variability for a sample of size $n = 500$ from a population of size 50,000,000.
- (c) about the same as the variability for a sample of size $n = 500$ from a population of size 50,000,000.
- (d) slightly greater than the variability for a sample of size $n = 500$ from a population of size 50,000,000.
- (e) much greater than the variability for a sample of size $n = 500$ from a population of size 50,000,000.

24. Increasing the sample size of an opinion poll will

- (a) reduce the bias of the poll result.
- (b) reduce the variability of the poll result.
- (c) reduce the effect of nonresponse on the poll.
- (d) reduce the variability of opinions.
- (e) all of the above.



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 on the next page.