CHAPTER 8

Estimation

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay."

Sherlock Holmes (A. C. Doyle), Adventures of the Copper Breeches

CONTENTS

8.1 Introduction 332

8.2 Point Estimator of a Population Mean 333

Problems 334

8.3 Point Estimator of a Population Proportion 336

Problems 342

8.4 Estimating a Population Variance 342

Problems 344

8.5 Interval Estimators of the Mean of a Normal Population with Known Population Variance 347

Problems 357

8.6 Interval Estimators of the Mean of a Normal Population with Unknown Population Variance 359

Problems 366

8.7 Interval Estimators of a Population Proportion 371

Problems 377

Key Terms 380

Summary 381

Review Problems 383

Introductory Statistics, DOI: 10.1016/B978-0-12-374388-6.00008-9

We learn how to use sample data to estimate a population mean, a population variance, and a population proportion. We discuss point estimates, which are single-value estimates of the parameter. The standard error of these estimates is considered. We also consider interval estimates that contain the parameter with specified degrees of confidence.

8.1 INTRODUCTION

It would not be unusual to see in a daily newspaper that "a recent poll of 1500 randomly chosen Americans indicates that 22 percent of the entire U.S. population is presently dieting, with a margin of error of ±2 percent." Perhaps you have wondered about such claims. For instance, what exactly does *with a margin of error of ±2 percent* mean? Also how is it possible, in a nation of over 150 million adults, that the proportion of them presently on diets can be ascertained by a sampling of only 1500 people?

In this chapter we will find the answers to these questions. In general, we will consider how one can learn about the numerical characteristics of a population by analyzing results from a sample of this population.

Whereas the numerical values of the members of the population can be summarized by a population probability distribution, this distribution is often not completely known. For instance, certain of its parameters, such as its mean and its standard deviation, may be unknown. A fundamental concern in statistics relates to how one can use the results from a sample of the population to estimate these unknown parameters.

For instance, if the items of the population consist of newly manufactured computer chips, then we may be interested in learning about the average functional lifetime of these chips. That is, we would be interested in estimating the population mean of the distribution of the lifetimes of these chips.

In this chapter we will consider ways of estimating certain parameters of the population distribution. To accomplish this, we will show how to use estimators and the estimates they give rise to.

Definition *An* estimator *is a statistic whose value depends on the particular sample drawn. The value of the estimator, called the* estimate, *is used to predict the value of a population parameter.*

For instance, if we want to estimate the mean lifetime of a chip, then we could employ the sample mean as an *estimator* of the population mean. If the value of the sample mean were 122 hours, then the *estimate* of the population mean would be 122 hours.

In Sec. 8.2 we consider the problem of estimating a population mean; in Sec. 8.3 we consider the problem of estimating a population proportion. Section 8.4

deals with estimating the population variance. The estimators considered in these sections are called *point estimators* because they are single values we hope will be close to the parameters they are estimating. In the remaining sections, we consider the problem of obtaining *interval estimators*. In this case, rather than specifying a particular value as our estimate, we specify an interval in which we predict that the parameter lies. We also consider the question of how much confidence to attach to a given interval estimate, that is, how certain we can be that the parameter indeed lies within this interval.

8.2 POINT ESTIMATOR OF A POPULATION MEAN

Let $X_1, \ldots, X_n$ denote a sample from a population whose mean μ is unknown. The sample mean $\overline{X}$ can be used as an estimator of μ. Since, as was noted in Sec. 7.3,

$$E[\overline{X}] = \mu$$

we see that the expected value of this estimator is the parameter we want to estimate. Such an estimator is called *unbiased*.

Definition *An estimator whose expected value is equal to the parameter it is estimating is said to be an* unbiased *estimator of that parameter.*

■ Example 8.1

To estimate the average amount of damages claimed in fires at medium-size apartment complexes, a consumer organization sampled the files of a large insurance company to come up with the following amounts (in thousands of dollars) for 10 claims:

$$121, 55, 63, 12, 8, 141, 42, 51, 66, 103$$

The estimate of the mean amount of damages claimed in all fires of the type being considered is thus

$$\begin{aligned}\overline{X} &= \frac{121 + 55 + 63 + 12 + 8 + 141 + 42 + 51 + 66 + 103}{10} \\ &= \frac{662}{10} = 66.2\end{aligned}$$

That is, we estimate that the mean fire damage claim is $66,200. ■

As we have shown, the sample mean $\overline{X}$ has expected value μ. Since a random variable is not likely to be too many standard deviations away from its expected value, it is important to determine the standard deviation of $\overline{X}$. However, as we have already noted in Sec. 7.3,

$$\mathrm{SD}(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

where σ is the population standard deviation. The quantity $SD(\overline{X})$ is sometimes called the *standard error* of $\overline{X}$ as an estimator of the mean. Since a random variable is unlikely to be more than 2 standard deviations away from its mean (especially when that random variable is approximately normal, as $\overline{X}$ will be when the sample size n is large), we are usually fairly confident that the estimate of the population mean will be correct to within ± 2 standard errors. Note that the standard error decreases by the square root of the sample size; as a result, to cut the standard error in half, we must increase the sample size by a factor of 4.

■ Example 8.2

Successive tests for the level of potassium in an individual's blood vary because of the basic imprecision of the test and because the actual level itself varies, depending on such things as the amount of food recently eaten and the amount of exertion recently undergone. Suppose it is known that, for a given individual, the successive readings of potassium level vary around a mean value μ with a standard deviation of 0.3. If a set of four readings on a particular individual yields the data

$$3.6, 3.9, 3.4, 3.5$$

then the estimate of the mean potassium level of that person is

$$\frac{3.6 + 3.9 + 3.4 + 3.5}{4} = 3.6$$

with the standard error of the estimate being equal to

$$SD(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{0.3}{2} = 0.15$$

Therefore, we can be quite confident that the actual mean will not differ from 3.6 by more than 0.30.

Suppose we wanted the estimator to have a standard error of 0.05. Then, since this would be a reduction in standard error by a factor of 3, it follows that we would have had to choose a sample 9 times as large. That is, we would have had to take 36 blood potassium readings. ■

PROBLEMS

1. The weights of a random sample of eight participants in the 2004 Boston Marathon were as follows:

$$121, 163, 144, 152, 186, 130, 128, 140$$

Use these data to estimate the average weight of all the participants in this race.

2. Suppose, in Prob. 1, that the data represented the weights of the top eight finishers in the marathon. Would you still be able to use these data to estimate the average weight of all the runners? Explain!

3. To determine the average amount of money spent by university students on textbooks, a random sample of 10 students was chosen, and the students were questioned. If the amounts (to the nearest dollar) spent were

   422, 146, 368, 52, 212, 454, 366, 711, 227, 680

   what is your estimate of the average amount spent by all students at the university?

4. A random sample of nine preschoolers from a given neighborhood yielded the following data concerning the number of hours per day each one spent watching television:

   3, 0, 5, 3.5, 1.5, 2, 3, 2.5, 2

   Estimate the average number of hours per day spent watching television by preschoolers in that neighborhood.

5. A manufacturer of compact disk players wants to estimate the average lifetime of the lasers in its product. A random sample of 40 is chosen. If the sum of the lifetimes of these lasers is 6624 hours, what is the estimate of the average lifetime of a laser?

6. A proposed study for estimating the average cholesterol level of working adults calls for a sample size of 1000. If we want to reduce the resulting standard error by a factor of 4, what sample size is necessary?

7. It is known that the standard deviation of the weight of a newborn child is 10 ounces. If we want to estimate the average weight of a newborn, how large a sample will be needed for the standard error of the estimate to be less than 3 ounces?

8. The following data represent the number of minutes each of a random sample of 12 recent patients at a medical clinic spent waiting to see a physician:

   46, 38, 22, 54, 60, 36, 44, 50, 35, 66, 48, 30

   Use these data to estimate the average waiting time of all patients at this clinic.

9. The following frequency table gives the household sizes of a random selection of 100 single-family households in a given city.

Household size	Frequency
1	11
2	19
3	28
4	26
5	11
6	4
7	1

Estimate the average size of all single-family households in the city.

10. Does (a) or (b) yield a more precise estimator of μ?
    (a) The sample mean of a sample of size n from a population with mean μ and variance σ^2
    (b) The sample mean of a sample of size $3n$ from a population with mean μ and variance $2\sigma^2$
    (c) How large would the sample in (b) have to be in order to match the precision of the estimator in (a)?
11. Repeat Prob. 10 when (a) and (b) are as follows:
    (a) The sample mean of a sample of size n from a population with mean μ and standard deviation σ
    (b) The sample mean of a sample of size $3n$ from a population with mean μ and standard deviation 3σ

8.3 POINT ESTIMATOR OF A POPULATION PROPORTION

Suppose that we are trying to estimate the proportion of a large population that is in favor of a given proposition. Let p denote the unknown proportion. To estimate p, a random sample should be chosen, and then p should be estimated by the proportion of the sample that is in favor. Calling this estimator $\hat{p}$, we can express it by

$$\hat{p} = \frac{X}{n}$$

where X is the number of members of the sample who are in favor of the proposition and n is the size of the sample.

From the results of Sec. 7.5, we know that

$$E[\hat{p}] = p$$

That is, $\hat{p}$, the proportion of the sample in favor of the proposition, is an unbiased estimator of p, the proportion of the entire population that is in favor. The spread of the estimator $\hat{p}$ about its mean p is measured by its standard deviation, which (again from Sec. 7.5) is equal to

$$\text{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

The standard deviation of $\hat{p}$ is also called the *standard error* of $\hat{p}$ as an estimator of the population proportion p. By the foregoing formula this standard error will be small whenever the sample size n is large. In fact, since it can be shown that for every value of p

$$p(1-p) \leq \frac{1}{4}$$

it follows that

$$\text{SD}(\hat{p}) \leq \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}$$

For instance, suppose a random sample of size 900 is chosen. Then no matter what proportion of the population is actually in favor of the proposition, it follows that the standard error of the estimator of this proportion is less than or equal to $1/(2\sqrt{900}) = 1/60$.

The preceding formula and bound on the standard error assume that we are drawing a random sample of size n from an infinitely large population. When the population size is smaller (as, of course, it will be in practice), then so is the standard error, thus making the estimator even more precise than just indicated.

■ Example 8.3

A school district is trying to determine its students' reaction to a proposed dress code. To do so, the school selected a random sample of 50 students and questioned them. If 20 were in favor of the proposal, then

(a) Estimate the proportion of all students who are in favor.
(b) Estimate the standard error of the estimate.

Solution

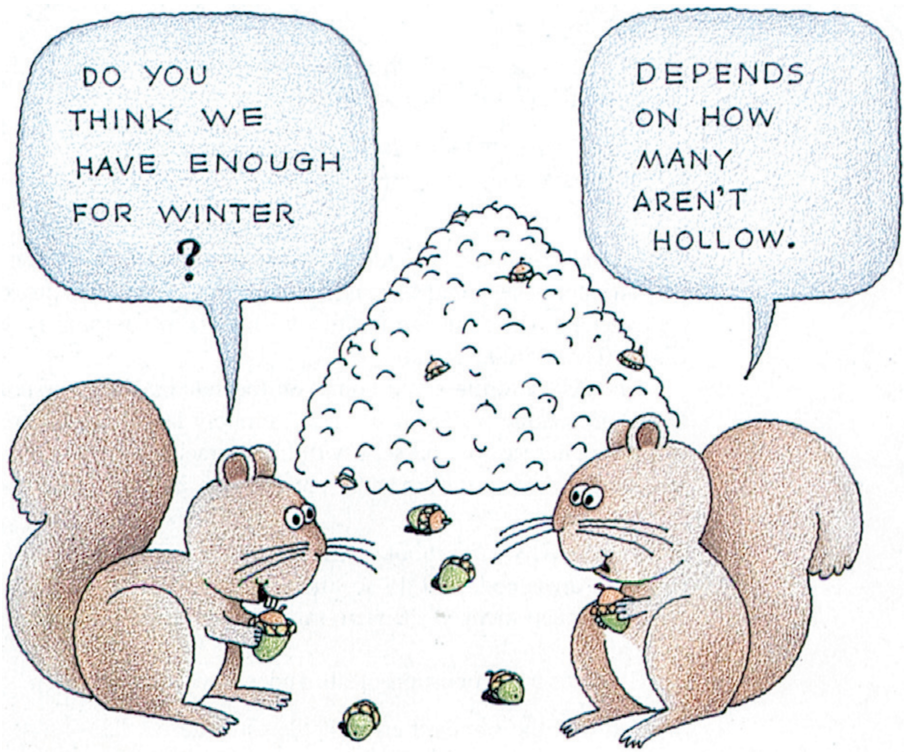
(a) The estimate of the proportion of all students who are in favor of the dress code is $20/50 = 0.40$.
(b) The standard error of the estimate is $\sqrt{p(1-p)/50}$, where p is the actual proportion of the entire population that is in favor. Using the estimate for p of 0.4, we can estimate this standard error by $\sqrt{0.4(1-0.4)/50} = 0.0693$. ■

PROBLEMS

1. In 1985, out of a random sample of 1325 North Americans questioned, 510 said that the Communist party would win a free election if one were held in the Soviet Union. Estimate the proportion of all North Americans who felt the same way at that time.
2. Estimate the standard error of the estimate in Prob. 1.
3. To learn the percentage of members who are in favor of increasing annual dues, a large social organization questioned a randomly chosen sample of 20 members. If 13 members were in favor, what is the estimate of the proportion of all members who are in favor? What is the estimate of the standard error?
4. The following are the results of 20 games of solitaire, a card game that results in either a win (w) or a loss (l):

$$w, l, l, l, w, l, l, w, l, w, w, l, l, l, l, w, l, l, w, l$$

   (a) Estimate the probability of winning a game of solitaire.
   (b) Estimate the standard error of the estimate in part (a).


DO YOU THINK WE HAVE ENOUGH FOR WINTER ?
DEPENDS ON HOW MANY AREN'T HOLLOW.

5. A random sample of 85 students at a large public university revealed that 35 students owned a car that was less than 5 years old. Estimate the proportion of all students at the university who own a car less than 5 years old. What is the estimate of the standard error of this estimate?

6. A random sample of 100 parents found that 64 are in favor of raising the driving age to 18.
   (a) Estimate the proportion of the entire population of parents who are in favor of raising the driving age to 18.
   (b) Estimate the standard error of the estimate in part (a).

7. A random sample of 1000 construction workers revealed that 122 are presently unemployed.
   (a) Estimate the proportion of all construction workers who are unemployed.
   (b) Estimate the standard error of the estimate in part (a).

8. Out of a random sample of 500 architects, 104 were women.
   (a) Estimate the proportion of all architects who are women.
   (b) Estimate the standard error of the estimate in part (a).

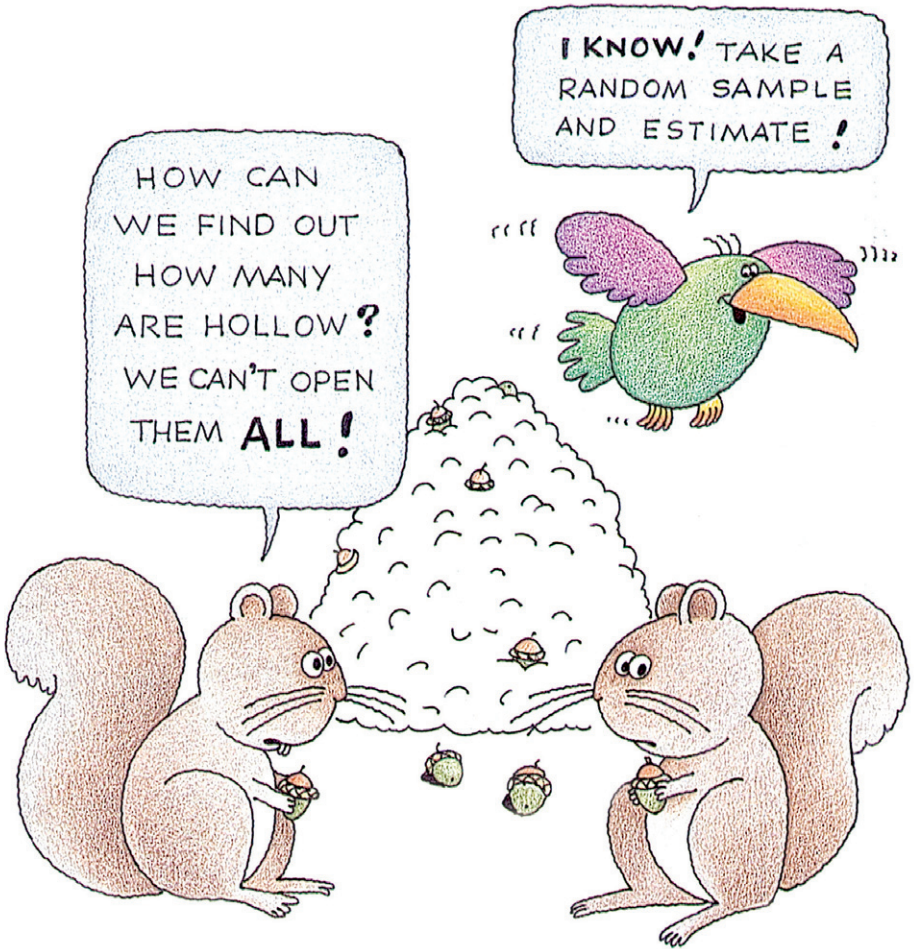
9. A random sample of 1200 engineers included 28 Hispanic Americans, 45 African Americans, and 104 females. Estimate the proportion of all engineers who are
   (a) Hispanic American
   (b) African American
   (c) Female

10. In parts (a), (b), and (c) of Prob. 9, estimate the standard error of the estimate.

11. A random sample of 400 death certificates related to teenagers yielded that 98 had died due to a motor vehicle accident.
   (a) Estimate the proportion of all teenage deaths due to motor vehicle accidents.
   (b) Estimate the standard error of the estimate in part (a).

12. A survey is being planned to discover the proportion of the population that is in favor of a new school bond. How large a sample is needed in order to be certain that the standard error of the resulting estimator is less than or equal to 0.1?

13. Los Angeles has roughly 3 times the voters of San Diego. Each city will be voting on a local education bond initiative. To determine the sentiments of the voters, a random sample of 3000 Los Angeles voters and a random sample of 1000 San Diego voters will be queried. Of the following statements, which is most accurate?
   (a) The resulting estimates of the proportions of people who will vote for the bonds in the two cities were equally accurate.


HOW CAN WE FIND OUT HOW MANY ARE HOLLOW? WE CAN'T OPEN THEM ALL!
I KNOW! TAKE A RANDOM SAMPLE AND ESTIMATE!

(b) The Los Angeles estimate is 3 times as accurate.
(c) The Los Angeles estimate is roughly 1.7 times as accurate.

Explain how you are interpreting the word *accurate* in statements (a), (b), and (c).

*14. The city of Chicago had 12,048 full-time law enforcement officers in 1990. To determine the number of African Americans in this group, a random sample of 600 officers was chosen, and it was discovered that 87 were African Americans.

(a) Estimate the number of African American law enforcement officers who were employed full-time in Chicago in 1990.
(b) Estimate the standard error of the estimate of part (a).

*8.3.1 Estimating the Probability of a Sensitive Event

Suppose that a company is interested in learning about the extent of illegal drug use among its employees. However, the company recognizes that employees might be reluctant to truthfully answer questions on this subject even if they have been assured that their answers will be kept in confidence. Indeed, even if the company assures workers that responses will not be traced to particular individuals, the employees might still remain suspicious and not answer truthfully. Given this background, how can the company elicit the desired information?

We now present a method that will enable the company to gather the desired information while at the same time protecting the privacy of those questioned. The method is to employ a randomization technique, and it works as follows: To begin, suppose that the sensitive question is stated in such a way that *yes* is the sensitive answer. For instance, the question could be, Have you used any illegal drugs in the past month? Presumably if the true answer is no, then the worker will not hesitate to give that answer. However, if the real answer is yes, then some workers may still answer no. To relieve any pressure to lie, the following rule for answering should be explained to each worker before the questioning begins: After the question has been posed, the worker is to flip a fair coin, not allowing the questioner to see the result of the flip. If the coin lands on heads, then the worker should answer yes to the question; and if it lands on tails, then the worker should answer the question honestly. It should be explained to the worker that an answer of yes does not mean that he or she is admitting to having used illegal drugs, since that answer may have resulted solely from the coin flip's landing on heads (which will occur 50 percent of the time). In this manner the workers sampled should feel assured that they can play the game truthfully and, at the same time, preserve their privacy.

Let us now analyze this situation to see how it can be used to estimate p, the proportion of the workforce that has actually used an illegal drug in the past month. Let $q = 1 - p$ denote the proportion that has not. Let us start by computing the probability that a sampled worker will answer no to the question. Since this will occur only if both (1) the coin toss lands on tails and (2) the worker has not used any illegal drugs in the past month, we see that

$$P\{\text{no}\} = \frac{1}{2} \times q = \frac{q}{2}$$

Hence, we can take the fraction of workers sampled who answered no as our estimate of $q/2$; or, equivalently, we can estimate q to be twice the proportion who answered no. Since $p = 1 - q$, this will also result in an estimate of p, the proportion of all workers who have used an illegal drug in the past month.

For instance, if 70 percent of the workers sampled answered the question in the affirmative, and so 30 percent answered no, then we would estimate that q was equal to $2(0.3) = 0.6$. That is, we would estimate that 60 percent of the population has not, and so 40 percent of the population has, used an illegal drug in the past month. If 35 percent of the workers answered no, then we would estimate that q was equal to $2(0.35) = 0.7$ and thus that $p = 0.3$. Similarly, if 48 percent of the workers answered no, then our estimate of p would be $1 - 2(0.48) = 0.04$.

Thus, by this trick of having each respondent flip a coin, we are able to obtain an estimate of p. However, the "price" we pay is an increased value of the standard error. Indeed, it can be shown that the standard error of the estimator of p is now $\sqrt{(1+p)(1-p)/n}$, which is larger than the standard error of the estimator when there is no need to use a coin flip (because all answers will be honestly given).

PROBLEMS

1. Suppose the randomization scheme described in this section is employed. If a sample of 50 people results in 32 yes answers, what is the estimate of p?
2. In Prob. 1, what would your estimate of p be if 40 of the 50 people answered yes?
3. When the randomization technique is used, the standard error of the estimator of p is $\sqrt{(1+p)(1-p)/n}$. Now, if there was no need to use the randomization technique, because everyone always answered honestly, then the standard error of the estimator of p would be $\sqrt{p(1-p)/n}$. The ratio of these standard errors is thus

$$\frac{\text{Standard error with randomization}}{\text{Usual standard error}} = \sqrt{\frac{1+p}{p}}$$

This ratio is thus an indicator of the price one must pay because of the sensitivity of the question.

(a) Do you think this price would be higher for large or small values of p?

(b) Determine the value of this ratio for $p = 0.1, 0.5$, and 0.9.

8.4 ESTIMATING A POPULATION VARIANCE

Suppose that we have a sample of size $n, X_1, \ldots, X_n$, from a population whose variance σ^2 is unknown, and that we are interested in using the sample data to estimate σ^2. The sample variance S^2, defined by

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

is an estimator of the population variance σ^2. To understand why, recall that the population variance is the expected squared difference between an observation and the population mean μ. That is, for $i = 1, \ldots, n$,

$$\sigma^2 = E[(X_i - \mu)^2]$$

Thus, it seems that the natural estimator of σ^2 would be the average of the squared differences between the data and the population mean μ. That is, it seems that the appropriate estimator of σ^2 would be

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}$$

This is indeed the appropriate estimator of σ^2 when the population mean μ is known. However, if the population mean μ is also unknown, then it is reasonable to use the foregoing expression with μ replaced by its estimator, namely, $\overline{X}$. To keep the estimator unbiased, this also leads us to change the denominator from n to $n - 1$; and thus we obtain the estimator S^2.

If the population mean μ is known, then the appropriate estimator of the population variance σ^2 is

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}$$

If the population mean μ is unknown, then the appropriate estimator of the population variance σ^2 is

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}$$

S^2 is an unbiased estimator of σ^2, that is,

$$E[S^2] = \sigma^2$$

Since the sample variance S^2 will be used to estimate the population variance σ^2, it is natural to use $\sqrt{S^2}$ to estimate the population standard deviation σ.

The population standard deviation σ is estimated by S, the sample standard deviation.

■ Example 8.4

A random sample of nine electronic components produced by a certain company yields the following sizes (in suitable units):

1211, 1224, 1197, 1208, 1220, 1216, 1213, 1198, 1197

What are the estimates of the population standard deviation and the population variance?

Solution

To answer this, we need to compute the sample variance S^2. Since subtracting a constant value from each data point will not affect the value of this statistic, start by subtracting 1200 from each datum to obtain the following transformed data set:

$$11, 24, -3, 8, 20, 16, 13, -2, -3$$

Using a calculator on these transformed data shows that the values of the sample variance and sample standard deviation are

$$S^2 = 103 \quad S = 10.149$$

Therefore, the respective estimates of the population standard deviation and the population variance are 10.149 and 103. ■

Statistics in Perspective

Variance Reduction Is the Key to Success in Manufacturing

According to Japanese quality control experts, the key to a successful manufacturing process—whether one is producing automobile parts, electronic equipment, computer chips, screws, or anything else—is to ensure that the production process consistently produces, at a reasonable cost, items that have values close to their *target* values. By this they mean that for any item being produced there is always a certain target value that the manufacturer is shooting at. For instance, when car doors are produced, there is a target value for the door's width. To be competitive, the widths of the doors produced must be consistently close to this value. These experts say that the key to producing items close to the target value is to ensure that the variance of the items produced is minimal. That is, once a production process has been established that produces items whose values have a small variance, then the difficult part of reaching the goal of consistently producing items whose values are near the target value has been accomplished.

Experience has led these experts to conclude that it is then a relatively simple matter to fine-tune the process so that the mean value of the item is close to the target value. (For an analogy, these experts are saying that if you want to build a rifle that will enable a shooter to consistently hit a particular target, then you should first concentrate on building a rifle that is extremely stable and will always give the same result when it is pointed in the same direction, and then you should train the shooter to shoot straight.)

PROBLEMS

1. A survey was undertaken to learn about the variation in the weekly number of hours worked by university professors. A sample of 10

professors yielded the following data:

48, 22, 19, 65, 72, 37, 55, 60, 49, 28

Use these data to estimate the population standard deviation of the number of hours that college professors work in a week.

2. The following data refer to the widths (in inches) of slots on nine successively produced duralumin forgings, which will be used as a terminal block at the end of an airplane wing span:

8.751, 8.744, 8.749, 8.750, 8.752, 8.749, 8.764, 8.746, 8.753

Estimate the mean and the standard deviation of the width of a slot.

3. The following data refer to the amounts (in tons) of chemicals produced daily at a chemical plant. Use them to estimate the mean and the variance of the daily production.

776, 810, 790, 788, 822, 806, 795, 807, 812, 791

4. Consistency is of great importance in manufacturing baseballs, for one does not want the balls to be either too lively or too dead. The balls are tested by dropping them from a standard height and then measuring how high they bounce. A sample of 30 balls resulted in the following summary statistics:

$$\sum_{i=1}^{30} X_i = 52.1 \qquad \sum_{i=1}^{30} X_i^2 = 136.2$$

Estimate the standard deviation of the size of the bounce. *Hint*: Recall the identity

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

5. Use the data of Prob. 1 of Sec. 8.2 to estimate the standard deviation of the weights of the runners in the 2004 Boston Marathon.

Problems 6, 7, and 8 refer to the following sample data:

104, 110, 114, 97, 105, 113, 106, 101, 100, 107

6. Estimate the population mean μ and the population variance σ^2.
7. Suppose it is known that the population mean is 104. Estimate the population variance.
8. Suppose it is known that the population mean is 106. Estimate the population standard deviation.

9. Use the data of Prob. 8 of Sec. 8.2 to estimate the standard deviation of the waiting times of patients at the medical clinic.

10. A manufacturer of furniture wants to test a sample of newly developed fire-resistant chairs to learn about the distribution of heat that these chairs can sustain before starting to burn. A sample of seven chairs is chosen, and each is put, one at a time, in a closed burn room. Once a chair is placed in this room, its temperature is increased, one degree at a time, until the chair bursts into flames. Suppose the burn temperatures for the seven chairs are (in degrees Fahrenheit) as follows:

    458, 440, 482, 455, 491, 477, 446

    (a) Estimate the mean burn temperature of this type of chair.
    (b) Estimate the standard deviation of the burn temperature of this type of chair.

11. Use the data of Prob. 9 of Sec. 8.2 to estimate the standard deviation of the size of a single-family household in the city considered.

12. Suppose that the systolic blood pressure of a worker in the mining industry is normally distributed. Suppose also that a random sample of 13 such workers yielded the following blood pressures:

    129, 134, 142, 114, 120, 116, 133, 142, 138, 148, 129, 133, 141

    (a) Estimate the mean systolic blood pressure of all miners.
    (b) Estimate the standard deviation of the systolic blood pressure.
    (c) Use the estimates in parts (a) and (b) along with the fact that the blood pressures are normally distributed to obtain an estimate of the proportion of all miners whose blood pressure exceeds 150.

13. The linear random walk model for the successive daily prices of a stock or commodity supposes that the successive differences of the end-of-day prices of a given stock constitute a random sample from a normal population. The following 20 data values represent the closing prices of crude oil on the New York Mercantile Exchange on 20 consecutive trading days in 1994. Assuming the linear random walk model, use these data to estimate the mean and standard deviation of the population distribution. (Note that the data give rise to 19 values from this distribution, the first being $17.60 - 17.50 = 0.10$, the second being $17.81 - 17.60 = 0.21$, and so on.)

    17.50, 17.60, 17.81, 17.67, 17.53, 17.39, 17.12, 16.71, 16.70, 16.83,
    17.21, 17.24, 17.22, 17.67, 17.83, 17.67, 17.55, 17.68, 17.98, 18.39

14. Due to a lack of precision in the scale used, the value obtained when a fish is weighed is normal with mean equal to the actual weight of

the fish and with standard deviation equal to 0.1 grams. A sample of 12 *different* fish was chosen, and the fish were weighed, with the following results:

5.5, 6.2, 5.8, 5.7, 6.0, 6.2, 5.9, 5.8, 6.1, 6.0, 5.7, 5.6

Estimate the population standard deviation of the actual weight of a fish.

Hint: First note that, due to the error involved in weighing a fish, each data value is not the true weight of a fish, but rather is the true weight plus an error term. This error term is an independent random variable that has mean 0 and standard deviation 0.1. Therefore,

$$\text{Data} = \text{true weight} + \text{error}$$

and so

$$\text{Var}\,(\text{data}) = \text{Var}\,(\text{true weight}) + \text{Var}\,(\text{error})$$

To determine the variance of the true weight, first estimate the variance of the data.

8.5 INTERVAL ESTIMATORS OF THE MEAN OF A NORMAL POPULATION WITH KNOWN POPULATION VARIANCE

When we estimate a parameter by a point estimator, we do not expect the resulting estimator to exactly equal the parameter, but we expect that it will be "close" to it. To be more specific, we sometimes try to find an interval about the point estimator in which we can be highly confident that the parameter lies. Such an interval is called an *interval estimator*.

Definition *An* interval estimator *of a population parameter is an interval that is predicted to contain the parameter. The* confidence *we ascribe to the interval is the probability that it will contain the parameter.*

To determine an interval estimator of a population parameter, we use the probability distribution of the point estimator of that parameter. Let us see how this works in the case of the interval estimator of a normal mean when the population standard deviation is assumed known.

Let $X_1, \ldots, X_n$ be a sample of size n from a normal population having known standard deviation σ, and suppose we want to utilize this sample to obtain a 95 percent confidence interval estimator for the population mean μ. To obtain such an interval, we start with the sample mean $\overline{X}$, which is the point estimator of μ. We now make use of the fact that $\overline{X}$ is normal with mean μ and standard deviation $\sigma/\sqrt{n}$, which implies that the standardized variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - \mu}{\sigma}$$

has a standard normal distribution. Now, since $z_{0.025} = 1.96$, it follows that 95 percent of the time the absolute value of Z is less than or equal to 1.96 (see Fig. 8.1).

Thus, we can write

$$P\left\{\frac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu\right| \leq 1.96\right\} = 0.95$$

Upon multiplying both sides of the inequality by $\sigma/\sqrt{n}$, we see that the preceding equation is equivalent to

$$P\left\{\left|\overline{X} - \mu\right| \leq 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

From the preceding statement we see that, with 95 percent probability, μ and $\overline{X}$ will be within $1.96\sigma/\sqrt{n}$ of each other. But this is equivalent to stating that

$$P\left\{\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

That is, with 95 percent probability, the interval $\overline{X} \pm 1.96\sigma/\sqrt{n}$ will contain the population mean.

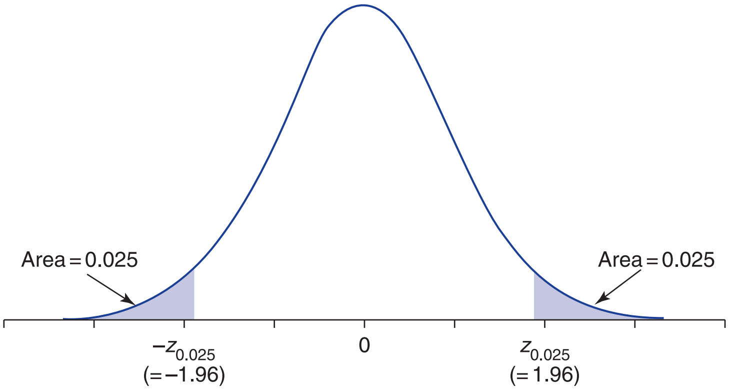


FIGURE 8.1
$P\{|Z| \leq 1.96\} = P\{-1.96 \leq Z \leq 1.96\} = 0.95.$

The interval from $\overline{X} - 1.96\sigma/\sqrt{n}$ to $\overline{X} + 1.96\sigma/\sqrt{n}$ is said to be a 95 *percent confidence interval estimator* of the population mean μ. If the observed value of $\overline{X}$ is $\overline{x}$, then we call the interval $\overline{x} \pm 1.96\sigma/\sqrt{n}$ a *95 percent confidence interval estimate* of μ.

In the long run, 95 percent of the interval estimates so constructed will contain the mean of the population from which the sample is drawn.

■ Example 8.5

Suppose that if a signal having intensity μ originates at location A, then the intensity recorded at location B is normally distributed with mean μ and standard deviation 3. That is, due to "noise," the intensity recorded differs from the actual intensity of the signal by an amount that is normal with mean 0 and standard deviation 3. To reduce the error, the same signal is independently recorded 10 times. If the successive recorded values are

$$17, 21, 20, 18, 19, 22, 20, 21, 16, 19$$

construct a 95 percent confidence interval for μ, the actual intensity.

Solution

The value of the sample mean is

$$\frac{17 + 21 + 20 + 18 + 19 + 22 + 20 + 21 + 16 + 19}{10} = 19.3$$

Since $\sigma = 3$, it follows that a 95 percent confidence interval estimate of μ is given by

$$19.3 \pm 19.6 \frac{3}{\sqrt{10}} = 19.3 \pm 1.86$$

That is, we can assert with 95 percent confidence that the actual intensity of the signal lies between 17.44 and 21.16. A picture of this confidence interval estimate is given in Fig. 8.2. ■

We can also consider confidence interval estimators having confidence levels different from 0.95. Recall that for any value of α between 0 and 1, the probability

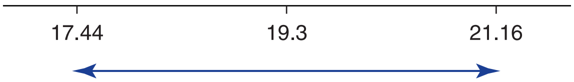


FIGURE 8.2

Confidence interval estimate of μ for Example 8.5.

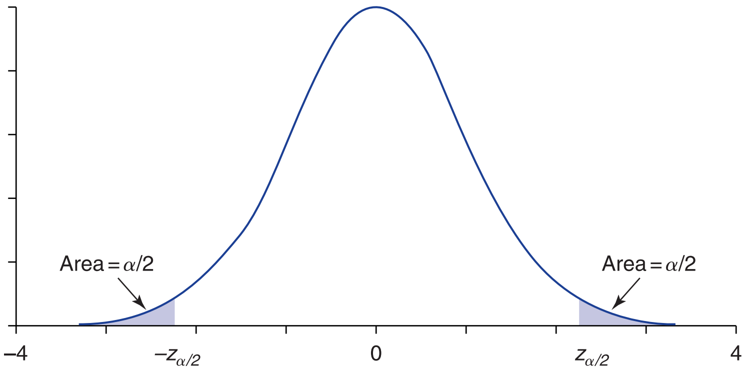


FIGURE 8.3

$P\{|Z| \le z_{\alpha/2}\} = P\{-z_{\alpha/2} \le Z \le z_{\alpha/2}\} = 1 - \alpha.$

Table 8.1 Confidence Level Percentiles

Confidence level $100(1-\alpha)$	Corresponding value of α	Value of $z_{\alpha/2}$
90	0.10	$z_{0.05} = 1.645$
95	0.05	$z_{0.025} = 1.960$
99	0.01	$z_{0.005} = 2.576$

that a standard normal lies in the interval between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is equal to $1-\alpha$ (Fig. 8.3). From this it follows that

$$P\left\{\frac{\sqrt{n}}{\sigma}\left|\overline{X}-\mu\right| \le z_{\alpha/2}\right\} = 1-\alpha$$

By the same logic used previously when $\alpha = 0.05(z_{0.025} = 1.96)$, we can show that, with probability $1-\alpha$, μ will lie in the interval $\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$.

The interval $\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$ is called a $100(1-\alpha)$ *percent confidence interval estimator* of the population mean.

Table 8.1 lists the values of $z_{\alpha/2}$ needed to construct 90, 95, and 99 percent confidence interval estimates of μ.

■ Example 8.6

Determine, for the data of Example 8.5,

(a) A 90 percent confidence interval estimate of μ
(b) A 99 percent confidence interval estimate of μ

Solution

We are being asked to construct a $100(1-\alpha)$ confidence interval estimate, with $\alpha = 0.10$ in part (a) and $\alpha = 0.01$ in part (b). Now

$$z_{0.05} = 1.645 \quad \text{and} \quad z_{0.005} = 2.576$$

and so the 90 percent confidence interval estimator is

$$\overline{X} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$

and the 99 percent confidence interval estimator is

$$\overline{X} \pm 2.576 \frac{\sigma}{\sqrt{n}}$$

For the data of Example 8.5, $n = 10, \overline{X} = 19.3$, and $\sigma = 3$. Therefore, the 90 and 99 percent confidence interval estimates for μ are, respectively,

$$19.3 \pm 1.645 \frac{3}{\sqrt{10}} = 19.3 \pm 1.56$$

and

$$19.3 \pm 2.576 \frac{3}{\sqrt{10}} = 19.3 \pm 2.44$$

Figure 8.4 indicates the 90, 95, and 99 percent confidence interval estimates of μ. Note that the larger the confidence coefficient $100(1-\alpha)$, the larger the length of this interval. This makes sense because if you want to increase your certainty that the parameter lies in a specified interval, then you will clearly have to enlarge that interval. ■

Sometimes we are interested in obtaining a $100(1-\alpha)$ percent confidence interval whose length is less than or equal to some specified value, and the problem

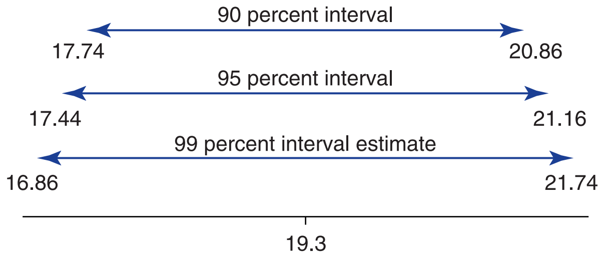


FIGURE 8.4

The 90, 95, and 99 percent confidence interval estimates.

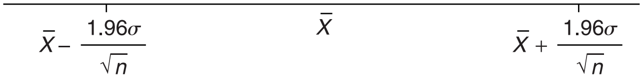

$\bar{X} - \frac{1.96\sigma}{\sqrt{n}}$
$\bar{X}$
$\bar{X} + \frac{1.96\sigma}{\sqrt{n}}$

FIGURE 8.5
The 95 percent confidence interval for μ.

is to choose the appropriate sample size. For instance, suppose we want to determine an interval of length at most b that, with 95 percent certainty, contains the population mean. How large a sample is needed? To answer this, note that since $z_{0.025} = 1.96$, a 95 percent confidence interval for μ based on a sample of size n is (see Fig. 8.5)

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

Since the length of this interval is

$$\text{Length of interval} = 2(1.96)\frac{\sigma}{\sqrt{n}} = 3.92\frac{\sigma}{\sqrt{n}}$$

we must choose n so that

$$\frac{3.92\sigma}{\sqrt{n}} \leq b$$

or, equivalently,

$$\sqrt{n} \geq \frac{3.92\sigma}{b}$$

Upon squaring both sides we see that the sample size n must be chosen so that

$$n \geq \left(\frac{3.92\sigma}{b}\right)^2$$

■ Example 8.7

If the population standard deviation is $\sigma = 2$ and we want a 95 percent confidence interval estimate of the mean μ that is of size less than or equal to $b = 0.01$, how large a sample is needed?

Solution

We have to select a sample of size n, where

$$n \geq \left(\frac{3.92 \times 2}{0.1}\right)^2 = (78.4)^2 = 6146.6$$

That is, a sample of size 6147 or larger is needed. ■

The analysis for determining the required sample size so that the length of a $100(1-\alpha)$ percent confidence interval is less than or equal to b is exactly the same as given when $\alpha = 0.05$. The result is as follows.

Determining the Necessary Sample Size

The length of the $100(1-\alpha)$ percent confidence interval estimator of the population mean will be less than or equal to b when the sample size n satisfies

$$n \geq \left(\frac{2z_{a/2}\sigma}{b}\right)^2$$

The confidence interval estimator is

$$\overline{X} \pm z_{a/2}\frac{\sigma}{\sqrt{n}}$$

■ Example 8.8

From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 90 percent certain that our estimate of the mean weight of a salmon is correct to within ±0.1 pounds, how large a sample is needed? What if we want to be 99 percent certain?

Solution

Since the 90 percent confidence interval estimator from a sample of size n will be $\overline{X} \pm 1.645\sigma/\sqrt{n}$, it follows that we can be 90 percent confident that the point estimator $\overline{X}$ will be within ± 0.1 of μ whenever the length of this confidence interval is less than or equal to 0.2 (see Fig. 8.6). Hence, from the preceding we see that n must be chosen so that

$$n \geq \left(\frac{2 \times 1.645 \times 0.3}{0.2}\right)^2 = 24.35$$

That is, a sample size of at least 25 is required.

On the other hand, if we wanted to be 99 percent certain that $\overline{X}$ will be within 0.1 pounds of the true mean, then since $z_{0.005} = 2.576$, the sample size n would

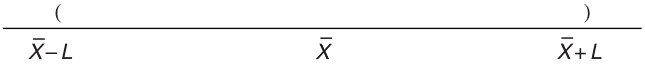


FIGURE 8.6

Confidence interval centered at $\overline{X}$. If length of interval is $2L$, then $\overline{X}$ is within L of any point in the interval.

need to satisfy

$$n \geq \left(\frac{2 \times 2.576 \times 0.3}{0.2}\right)^2 = 59.72$$

That is, a sample of size 60 or more is needed. ■

In deriving confidence interval estimators of a normal mean whose variance σ^2 is known, we used the fact that $\overline{X}$ is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. However, by the central limit theorem, this will remain approximately true for the sample mean of any population distribution provided the sample size n is relatively large ($n \geq 30$ is almost always sufficiently large). As a result, we can use the interval $\overline{X} \pm z_{a/2}\sigma/\sqrt{n}$ as a $100(1-\alpha)$ percent confidence interval estimator of the population mean for any population provided the sample size is large enough for the central limit theorem to apply.

■ Example 8.9

To estimate μ, the average nicotine content of a newly marketed cigarette, 44 of these cigarettes are randomly chosen, and their nicotine contents are determined.

(a) If the average nicotine finding is 1.74 milligrams, what is a 95 percent confidence interval estimator of μ?

(b) How large a sample is necessary for the length of the 95 percent confidence interval to be less than or equal to 0.3 milligrams?

Assume that it is known from past experience that the standard deviation of the nicotine content of a cigarette is equal to 0.7 milligrams.

Solution

(a) Since 44 is a large sample size, we do not have to suppose that the population distribution is normal to assert that a 95 percent confidence interval estimator of the population mean is

$$\overline{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}}$$

In this case, the estimator reduces to

$$1.74 \pm \frac{1.96(0.7)}{\sqrt{44}} = 1.74 \pm 0.207$$

That is, we can assert with 95 percent confidence that the average amount of nicotine per cigarette lies between 1.533 and 1.947 milligrams.

(b) The length of the 95 percent confidence interval estimate will be less than or equal to 0.3 if the sample size n is large enough that

$$n \geq \left(\frac{2 \times 1.96 \times 0.7}{0.3}\right)^2 = 83.7$$

That is, a sample size of at least 84 is needed. ■

8.5.1 Lower and Upper Confidence Bounds

Sometimes we are interested in making a statement to the effect that a population mean is, with a given degree of confidence, greater than some stated value. To obtain such a *lower confidence bound* for the population mean, we again use the fact that

$$Z = \sqrt{n}\frac{\overline{X} - \mu}{\sigma}$$

has a standard normal distribution. As a result, it follows that (see Fig. 8.7)

$$P\left\{\sqrt{n}\frac{\overline{X} - \mu}{\sigma} < z_\alpha\right\} = 1 - \alpha$$

which can be rewritten as

$$P\left\{\mu > \overline{X} - z_\alpha\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

From this equation, we can conclude the following.

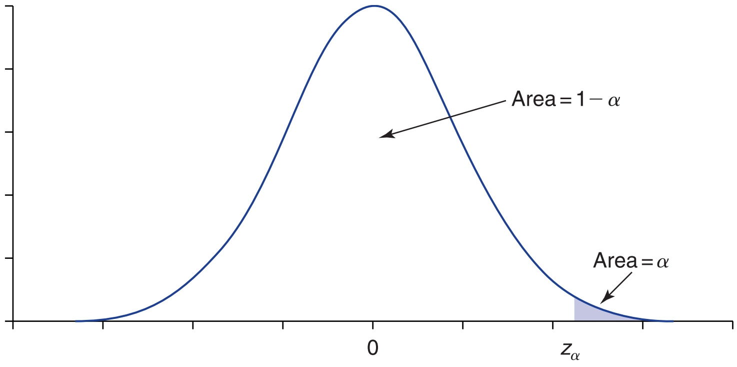


FIGURE 8.7

$P\{Z \leq z_\alpha\} = 1 - \alpha$.

A $100(1-\alpha)$ percent lower confidence bound for the population mean μ is given by

$$\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

That is, with $100(1-\alpha)$ percent confidence, we can assert that

$$\mu > \overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

■ Example 8.10

Suppose in Example 8.8 that we want to specify a value that, with 95 percent confidence, is less than the average weight of a salmon. If a sample of 50 salmon yields an average weight of 5.6 pounds, determine this value.

Solution

We are asked to find a 95 percent lower confidence bound for μ. By the preceding analysis this will be given by

$$\overline{X} - z_{0.05} \frac{\sigma}{\sqrt{n}}$$

Since $z_{0.05} = 1.645$, $\sigma = 0.3$, $n = 50$, and $\overline{X} = 5.6$, the lower confidence bound will equal

$$5.6 - 1.645 \frac{0.3}{\sqrt{50}} = 5.530$$

That is, we can assert, with 95 percent confidence, that the mean weight of a salmon is greater than 5.530 pounds. ■

We can also derive a $100(1-\alpha)$ percent upper confidence bound for μ. The result is the following.

A $100(1-\alpha)$ percent upper confidence bound for the population mean μ is given by

$$\overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

That is, with $100(1-\alpha)$ percent confidence, we can assert that

$$\mu < \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

■ Example 8.11

In Example 8.9, find a 95 percent upper confidence bound for μ.

Solution

A 95 percent upper confidence bound is given by

$$\overline{X} + z_{0.05}\frac{\sigma}{\sqrt{n}} = 1.74 + 1.645\frac{0.7}{\sqrt{44}} = 1.914$$

That is, we can assert with 95 percent confidence that the average nicotine content is less than 1.914 milligrams. ■

PROBLEMS

1. An electric scale gives a reading equal to the true weight plus a random error that is normally distributed with mean 0 and standard deviation $\sigma = 0.1$ ounces. Suppose that the results of five successive weighings of the same object are as follows: 3.142, 3.163, 3.155, 3.150, 3.141.
   (a) Determine a 95 percent confidence interval estimate of the true weight.
   (b) Determine a 99 percent confidence interval estimate of the true weight.
2. Suppose that a hospital administrator states that a statistical experiment has indicated that “With 90 percent certainty, the average weight at birth of all boys born at the certain hospital is between 6.6 and 7.2 pounds.” How would you interpret this statement?
3. The polychlorinated biphenyl (PCB) concentration of a fish caught in Lake Michigan was measured by a technique that is known to result in an error of measurement that is normally distributed with standard deviation 0.08 parts per million. If the results of 10 independent measurements of this fish are

   11.2, 12.4, 10.8, 11.6, 12.5, 10.1, 11.0, 12.2, 12.4, 10.6

   give a 95 percent confidence interval estimate of the PCB level of this fish.
4. Suppose in Prob. 3 that 40 measurements are taken, with the same average value resulting as in Prob. 3. Again determine a 95 percent confidence interval estimate of the PCB level of the fish tested.
5. The life of a particular brand of television picture tube is known to be normally distributed with a standard deviation of 400 hours. Suppose

that a random sample of 20 tubes resulted in an average lifetime of 9000 hours. Obtain a

(a) 90 percent
(b) 95 percent

confidence interval estimate of the mean lifetime of such a tube.

6. An engineering firm manufactures a space rocket component that will function for a length of time that is normally distributed with a standard deviation of 3.4 hours. If a random sample of nine such components has an average life of 10.8 hours, find a

(a) 95 percent
(b) 99 percent

confidence interval estimate of the mean length of time that these components function.

7. The standard deviation of test scores on a certain achievement test is 11.3. A random sample of 81 students had a sample mean score of 74.6. Find a 90 percent confidence interval estimate for the average score of all students.

8. In Prob. 7, suppose the sample mean score was 74.6 but the sample was of size 324. Again find a 90 percent confidence interval estimate.

9. The standard deviation of the lifetime of a certain type of lightbulb is known to equal 100 hours. A sample of 169 such bulbs had an average life of 1350 hours. Find a

(a) 90 percent
(b) 95 percent
(c) 99 percent

confidence interval estimate of the mean life of this type of bulb.

10. The average life of a sample of 10 tires of a certain brand was 28,400 miles. If it is known that the lifetimes of such tires are normally distributed with a standard deviation of 3300 miles, determine a 95 percent confidence interval estimate of the mean life.

11. For Prob. 10, how large a sample would be needed to obtain a 99 percent confidence interval estimator of smaller size than the interval obtained in the problem?

12. A pilot study has revealed that the standard deviation of workers' monthly earnings in the chemical industry is $180. How large a sample must be chosen to obtain an estimator of the mean salary that, with 90 percent confidence, will be correct to within ±$20?

13. Repeat Prob. 12 for when you require 95 percent confidence.

14. A college admissions officer wanted to know the average Scholastic Aptitude Test (SAT) score of this year's class of entering students. Rather than checking all student folders, she decided to use a randomly chosen sample. If it is known that student scores are normally distributed with a standard deviation of 70, how large a random

sample is needed if the admissions officer wants to obtain a 95 percent confidence interval estimate that is of length 4 or less?

15. In Prob. 7, find a
   (a) 90 percent lower confidence bound
   (b) 95 percent lower confidence bound
   (c) 95 percent upper confidence bound
   (d) 99 percent upper confidence bound

   for the average test score.

16. The following are data from a normal population with standard deviation 3:

$$5, 4, 8, 12, 11, 7, 14, 12, 15, 10$$

   (a) Find a value that, with 95 percent confidence, is larger than the population mean.
   (b) Find a value that, with 99 percent confidence, is smaller than the population mean.

17. Suppose, on the basis of the sample data noted in Prob. 10, that the tire manufacturer advertises, "With 95 percent certainty, the average tire life is over 26,000 miles." Is this false advertising?

8.6 INTERVAL ESTIMATORS OF THE MEAN OF A NORMAL POPULATION WITH UNKNOWN POPULATION VARIANCE

Suppose now that we have a sample $X_1, \ldots, X_n$ from a normal population having an unknown mean μ and an unknown standard deviation σ, and we want to use the sample data to obtain an interval estimator of the population mean μ.

To start, let us recall how we obtained the interval estimator of μ when σ was assumed to be known. This was accomplished by using the fact that Z, the standardized version of the point estimator $\overline{X}$, which is given by

$$Z = \sqrt{n}\frac{\overline{X} - \mu}{\sigma}$$

has a standard normal distribution. Since σ is no longer known, it is natural to replace it by its estimator S, the sample standard deviation, and thus to base our confidence interval on the variable T_{n-1} given by

$$T_{n-1} = \sqrt{n}\frac{\overline{X} - \mu}{S}$$

The random variable T_{n-1} just defined is said to be a t random variable having $n - 1$ degrees of freedom.

The random variable

$$T_{n-1} = \sqrt{n}\frac{\overline{X} - \mu}{S}$$

is said to be a *t random variable having* $n - 1$ *degrees of freedom*.

The reason that T_{n-1} has $n - 1$ degrees of freedom is that the sample variance S^2, which is being used to estimate σ^2, has, when multiplied by $(n - 1)/\sigma^2$, a chi-squared distribution with $n - 1$ degrees of freedom (see Sec. 7.6).

The density function of a t random variable, like a standard normal random variable, is symmetric about zero. It looks similar to a standard normal density, although it is somewhat more spread out, resulting in its having "larger tails." As the degree of freedom parameter increases, the density becomes more and more similar to the standard normal density. Figure 8.8 depicts the probability density functions of t random variables for a variety of different degrees of freedom.

The quantity $t_{n,\alpha}$ is defined to be such that

$$P\{T_n > t_{n,\alpha}\} = \alpha$$

where T_n is a t random variable with n degrees of freedom (see Fig. 8.9).

Since $P\{T_n < t_{n,\alpha}\} = 1 - \alpha$, it follows that $t_{n,\alpha}$ is the $100(1 - \alpha)$ percentile of the t distribution with n degrees of freedom. For instance, $P\{T_n < t_{n,0.05}\} = 0.095$,

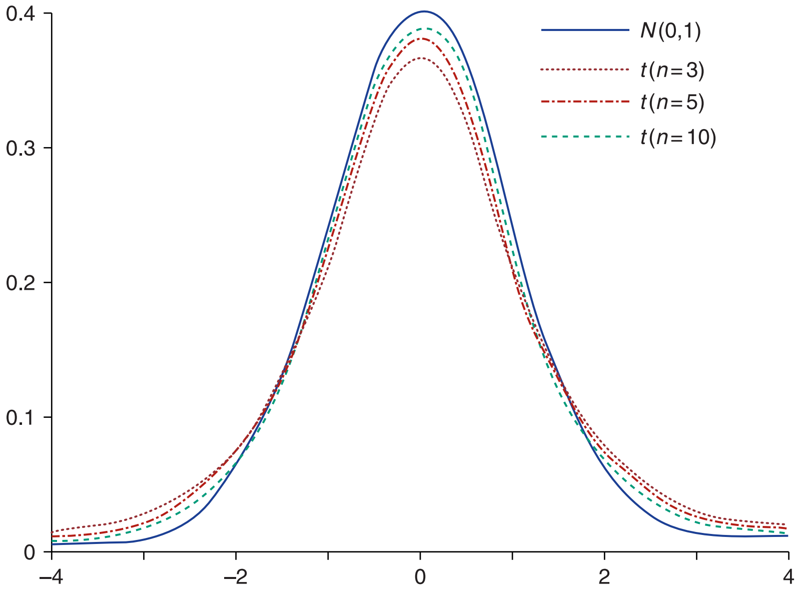


FIGURE 8.8

Standard normal and t distributions.

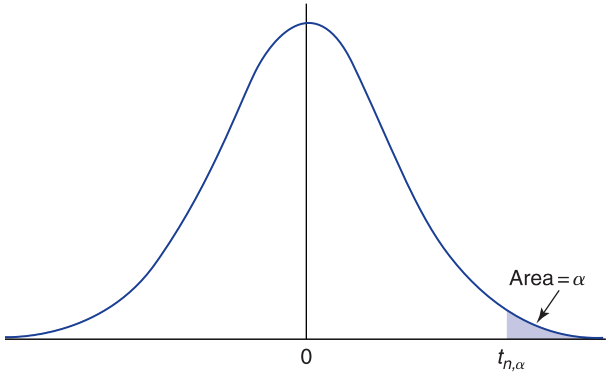


FIGURE 8.9
The t density percentile: $P\{T_n > t_{n,\alpha}\} = \alpha$.

showing that 95 percent of the time a t random variable having n degrees of freedom will be less than $t_{n,0.05}$. The quantity $t_{n,\alpha}$ is analogous to the quantity z_α of the standard normal distribution.

Values of $t_{n,\alpha}$ for various values of n and α are presented in App. D, Table D.2. In addition, Program 8-1 will compute the value of these percentiles. Program 8-2 can also be used to compute the probabilities of a t random variable.

Example 8.12

Find $t_{8,0.05}$.

Solution

The value of $t_{8,0.05}$ can be obtained from Table D.2. The following is taken from that table.

Values of $t_{n,\alpha}$

n	$\alpha = 0.10$	↓ $\alpha = 0.05$	$\alpha = 0.025$
6	1.440	1.943	2.447
7	1.415	1.895	2.365
→ 8	1.397	1.860	2.306
9	1.383	1.833	2.262

Reading down the $\alpha = 0.05$ column for the row $n = 8$ shows that $t_{8,0.05} = 1.860$.

By the symmetry of the t distribution about zero, it follows (see Fig. 8.10) that

$$P\left\{|T_n| \leq t_{n,\alpha/2}\right\} = P\left\{-t_{n,\alpha/2} \leq T_n \leq t_{n,\alpha/2}\right\} = 1 - \alpha$$

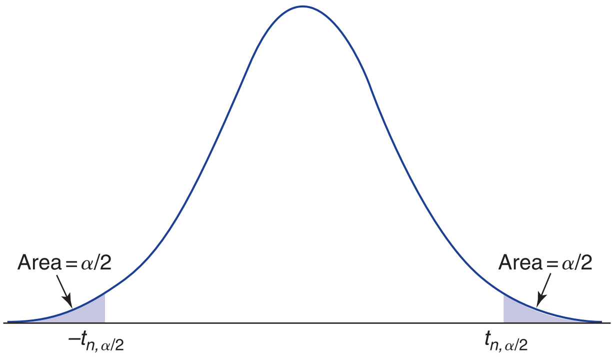


FIGURE 8.10

$P\{|T_n| \leq t_{n,\alpha/2}\} = P\{-t_{n,\alpha/2} \leq T_n \leq t_{n,\alpha/2}\} = 1 - \alpha.$

Hence, upon using the result that $\sqrt{n}\left(\overline{X} - \mu\right)/S$ has a t distribution with $n - 1$ degrees of freedom, we see that

$$P\left\{\sqrt{n}\frac{|\overline{X} - \mu|}{S} \leq t_{n-1,\alpha/2}\right\} = 1 - \alpha$$

In exactly the same manner as we did when σ was known, we can show that the preceding equation is equivalent to

$$P\left\{\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right\} = 1 - \alpha$$

Therefore, we showed the following.

A $100(1 - \sigma)$ percent confidence interval estimator for the population mean μ is given by the interval

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

Program 8-3 will compute the desired confidence interval estimate for a given data set.

■ Example 8.13

The Environmental Protection Agency (EPA) is concerned about the amounts of PCB, a toxic chemical, in the milk of nursing mothers. In a sample of 20 women, the amounts (in parts per million) of PCB were as follows:

$$16, 0, 0, 2, 3, 6, 8, 2, 5, 0, 12, 10, 5, 7, 2, 3, 8, 17, 9, 1$$

Use these data to obtain a

(a) 95 percent confidence interval
(b) 99 percent confidence interval

of the average amount of PCB in the milk of nursing mothers.

Solution

A simple calculation yields that the sample mean and sample standard deviation are

$$\overline{X} = 5.8 \qquad S = 5.085$$

Since $100(1-\alpha)$ equals 0.95 when $\alpha = 0.05$ and equals 0.99 when $\alpha = 0.01$, we need the values of $t_{19,0.025}$ and $t_{19,0.005}$. From Table D.2 we see that

$$t_{19,0.025} = 2.093 \qquad t_{19,0.005} = 2.861$$

Hence, the 95 percent confidence interval estimate of μ is

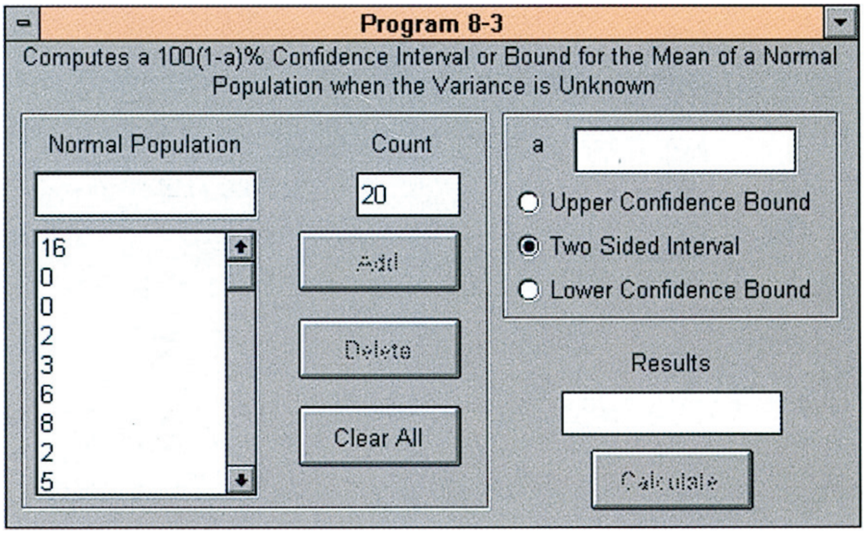
$$5.8 \pm 2.093 \frac{5.085}{\sqrt{20}} = 5.8 \pm 2.38$$

and the 99 percent confidence interval estimate of μ is

$$5.8 \pm 2.861 \frac{5.085}{\sqrt{20}} = 5.8 \pm 3.25$$

That is, we can be 95 percent confident that the average amount of PCB in the milk of nursing mothers is between 3.42 and 8.18 parts per million; and we can be 99 percent confident that it is between 2.55 and 9.05 parts per million.

This could also have been solved by running Program 8-3, which yields the following.



Program 8-3

Computes a 100(1-a)% Confidence Interval or Bound for the Mean of a Normal Population when the Variance is Unknown

Normal Population

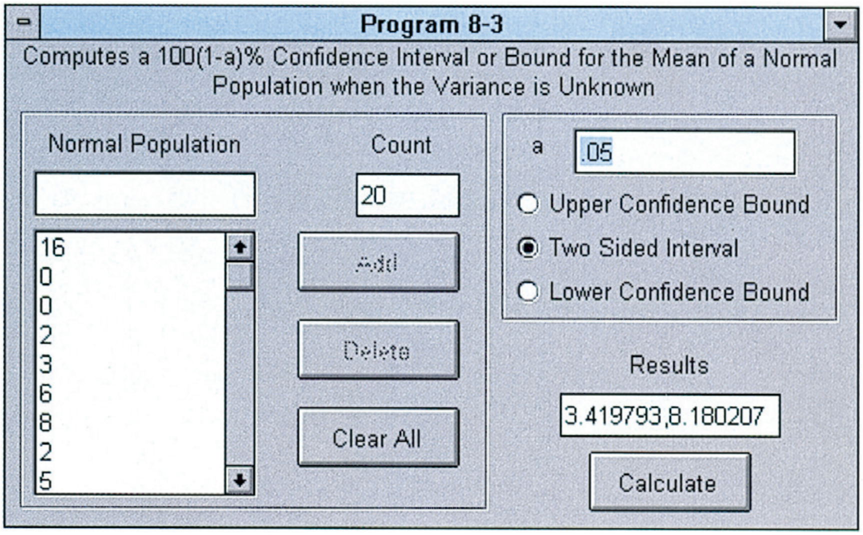
Count: 20

16
0
0
2
3
6
8
2
5

a

○ Upper Confidence Bound
● Two Sided Interval
○ Lower Confidence Bound

Results


Program 8-3

Computes a 100(1-a)% Confidence Interval or Bound for the Mean of a Normal Population when the Variance is Unknown

Normal Population

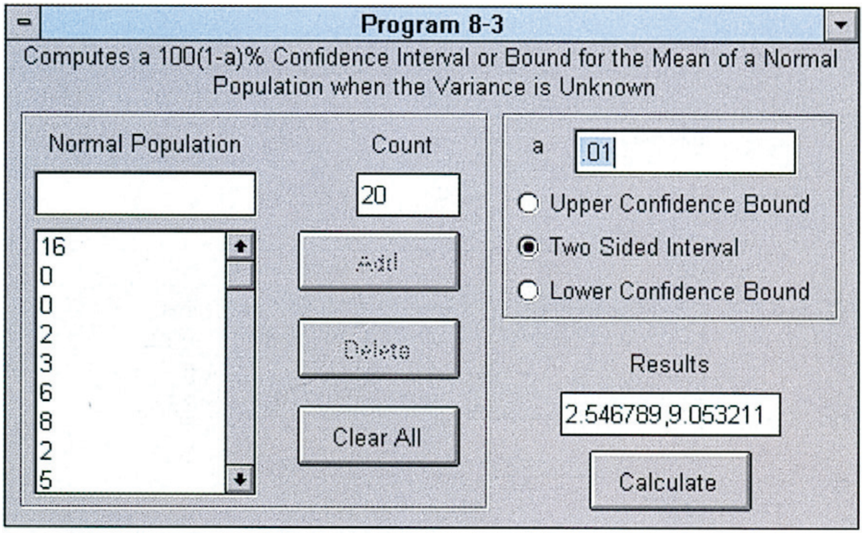
Count: 20

16
0
0
2
3
6
8
2
5

a: .05

○ Upper Confidence Bound
● Two Sided Interval
○ Lower Confidence Bound

Results: 3.419793,8.180207


Program 8-3

Computes a 100(1-a)% Confidence Interval or Bound for the Mean of a Normal Population when the Variance is Unknown

Normal Population

Count: 20

16
0
0
2
3
6
8
2
5

a: .01

○ Upper Confidence Bound
● Two Sided Interval
○ Lower Confidence Bound

Results: 2.546789,9.053211

8.6.1 Lower and Upper Confidence Bounds

Lower and upper confidence bounds for μ are also easily derived, with the following results.

A $100(1-\alpha)$ percent lower confidence bound for μ is given by

$$\overline{X} - t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

That is, with $100(1-\alpha)$ percent confidence the population mean is greater than

$$\overline{X} - t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

A $100(1-\alpha)$ percent upper confidence bound for μ is given by

$$\overline{X} + t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

That is, with $100(1-\alpha)$ percent confidence the population mean is less than

$$\overline{X} + t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

■ Example 8.14

In Example 8.13, find a

(a) 95 percent upper confidence bound
(b) 99 percent lower confidence bound

for the average amount of PCB in nursing mothers.

Solution

The sample size in Example 8.13 was equal to 20, and the values of the sample mean and the sample standard deviation were

$$\overline{X} = 5.8 \qquad S = 5.085$$

(a) From Table D.2 we see that

$$t_{19,0.05} = 1.729$$

Therefore, the 95 percent upper confidence bound is

$$5.8 + 1.729\frac{5.085}{\sqrt{20}} = 7.77$$

That is, we can be 95 percent confident that the average PCB level in the milk of nursing mothers is less than 7.77 parts per million.

(b) From Table D.2,

$$t_{19,0.01} = 2.539$$

Therefore, the 99 percent lower confidence bound is

$$5.8 - 2.539\frac{5.085}{\sqrt{20}} = 2.91$$

and so we can be 99 percent confident that the average PCB level in the milk of nursing mothers is greater than 2.91 parts per million. ■

Program 8-3 will also compute upper and lower confidence bounds having any desired confidence level.

PROBLEMS

1. The National Center for Educational Statistics recently chose a random sample of 2000 newly graduated college students and queried each one about the time it took to complete his or her degree. If the sample mean was 5.2 years with a sample standard deviation of 1.2 years, construct
   (a) A 95 percent confidence interval estimate of the mean completion time of all newly graduated students
   (b) A 99 percent confidence interval estimate
2. The manager of a shipping department of a mail-order operation located in New York has been receiving complaints about the length of time it takes for customers in California to receive their orders. To learn more about this potential problem, the manager chose a random sample of 12 orders and then checked to see how many days it took to receive each of these orders. The resulting data were

   15, 20, 10, 11, 7, 12, 9, 14, 12, 8, 13, 16

   (a) Find a 90 percent confidence interval estimate for the mean time it takes California customers to receive their orders.
   (b) Find a 95 percent confidence interval estimate.
3. A survey was instituted to estimate μ, the mean salary of middle-level bank executives. A random sample of 15 executives yielded the following yearly salaries (in units of $1000):

   88, 121, 75, 39, 52, 102, 95, 78, 69, 82, 80, 84, 72, 115, 106

   Find a
   (a) 90 percent
   (b) 95 percent
   (c) 95 percent
   confidence interval estimate of μ.

4. The numbers of riders on an intercity bus on 12 randomly chosen days are

$$47, 66, 55, 53, 49, 65, 48, 44, 50, 61, 60, 55$$

   (a) Estimate the mean number of daily riders.
   (b) Estimate the standard deviation of the daily number of riders.
   (c) Give a 95 percent confidence interval estimate for the mean number of daily riders.

5. Use the data of Prob. 1 of Sec. 8.2 to obtain a
   (a) 95 percent
   (b) 99 percent

   confidence interval estimate of the average weight of all participants of the 2004 Boston Marathon.

6. A random sample of 30 General Electric transistors resulted in an average lifetime of 1210 hours with a sample standard deviation of 92 hours. Compute a
   (a) 90 percent
   (b) 95 percent
   (c) 99 percent

   confidence interval estimate of the mean life of all General Electric transistors.

7. In Prob. 10 of Sec. 8.4 determine a 95 percent confidence interval estimate of the population mean burn temperature.

8. The following are the losing scores in seven randomly chosen Super Bowl football games:

$$10, 16, 20, 17, 31, 19, 14$$

   Construct a 95 percent confidence interval estimate of the average losing score in a Super Bowl game.

9. The following are the winning scores in eight randomly chosen Masters Golf Tournaments:

$$285, 279, 280, 288, 279, 286, 284, 279$$

   Use these data to construct a 90 percent confidence interval estimate of the average winning score in the Masters.

10. All the students at a certain school are to be given a psychological task. To determine the average time it will take a student to perform this task, a random sample of 20 students was chosen and each was given the task. If it took these students an average of 12.4 minutes to complete the task with a sample standard deviation of 3.3 minutes, find a 95 percent confidence interval estimate for the average time it will take all students in the school to perform this task.

11. A large company self-insures its large fleet of cars against collisions. To determine its mean repair cost per collision, it has randomly chosen a sample of 16 accidents. If the average repair cost in these accidents is $2200 with a sample standard deviation of $800, find a 90 percent confidence interval estimate of the mean cost per collision.

12. An anthropologist measured the heights (in inches) of a random sample of 64 men of a certain tribe, and she found that the sample mean was 72.4 and the sample standard deviation was 2.2. Find a

    (a) 95 percent
    (b) 99 percent

    confidence interval estimate of the average height of all men of the tribe.

13. To determine the average time span of a phone call made during midday, the telephone company has randomly selected a sample of 1200 such calls. The sample mean of these calls is 4.7 minutes, and the sample standard deviation is 2.2 minutes. Find a

    (a) 90 percent
    (b) 95 percent

    confidence interval estimate of the mean length of all such calls.

14. Each of 20 science students independently measured the melting point of lead. The sample mean and sample standard deviation of these measurements were (in degrees Celsius) 330.2 and 15.4, respectively. Construct a

    (a) 95 percent
    (b) 99 percent

    confidence interval estimate of the true melting point of lead.

15. A random sample of 300 Citibank VISA cardholder accounts indicated a sample mean debt of $1220 with a sample standard deviation of $840. Construct a 95 percent confidence interval estimate of the average debt of all cardholders.

16. To obtain information about the number of years that Chicago police officers have been on the job, a sample of 46 officers was chosen. Their average time on the job was 14.8 years with a sample standard deviation of 8.2 years. Determine a

    (a) 90 percent
    (b) 95 percent
    (c) 99 percent

    confidence interval estimate for the average time on the job of all Chicago police officers.

17. The following statement was made by an "expert" in statistics. "If a sample of size 9 is chosen from a normal distribution having mean μ, then we can be 95 percent certain that μ will lie within $\overline{X} \pm 1.96S/3$

where $\overline{X}$ is the sample mean and S is the sample standard deviation.” Is this statement correct?

18. The geometric random walk model for the price of a stock supposes that the successive differences in the logarithms of the closing prices of the stock constitute a sample from a normal population. This implies that the percentage changes in the successive closing prices constitute a random sample from a population (as opposed to the linear random walk model given in Problem 13 of Sec. 8.4, which supposes that the magnitudes of the changes constitute a random sample). Thus, for instance, under the geometric random walk model, the chance that a stock whose price is 100 will increase to 102 is the same as the chance when its price is 50 that it will increase to 51.

The following data give the logarithms and the successive differences of the logarithms of the closing crude oil prices of 20 consecutive trading days in 1994. Assuming the applicability of the geometric random walk model, use them to construct a 95 percent confidence interval for the population mean.

Price	log (price)	log (price) difference
17.50	2.862201	
17.60	2.867899	5.697966E – 03
17.81	2.87976	1.186109E – 02
17.67	2.871868	–7.891655E – 03
17.53	2.863914	–7.954597E – 03
17.39	2.855895	–8.018494E – 03
17.12	2.840247	–1.564789E – 02
16.71	2.816007	–2.424002E – 02
16.70	2.815409	–5.986691E – 04
16.83	2.823163	7.754326E – 03
17.21	2.84549	2.232742E – 02
17.24	2.847232	1.741886E – 03
17.22	2.846071	–1.16086E – 03
17.67	2.871868	2.579689E – 02
17.83	2.880883	9.01413E – 03
17.67	2.871868	–9.01413E – 03
17.55	2.865054	–6.81448E – 03
17.68	2.872434	7.380247E – 03
17.98	2.88926	1.682591E – 02
18.39	2.911807	2.254701E – 02

19. Twelve successively tested lightbulbs functioned for the following lengths of time (measured in hours):

35.6, 39.2, 18.4, 42.0, 45.3, 34.5, 27.9, 24.4, 19.9, 40.1, 37.2, 32.9

(a) Give a 95 percent confidence interval estimate of the mean life of a lightbulb.

(b) A claim has been made that the results of this experiment indicate that "One can be 99 percent certain that the mean life exceeds 30 hours." Do you agree with this statement?

20. A school principal was instructed by his board to determine the average number of school days missed by students in the past year. Rather than making a complete survey of all students, the principal drew a random sample of 50 names. He then discovered that the average number of days missed by these 50 students was 8.4 with a sample standard deviation of 5.1.

(a) What is a 95 percent confidence interval estimate of the average number of days missed by all students?

(b) At a subsequent board meeting the principal stated, "With 95 percent confidence I can state that the average number of days missed is less than ______." Fill in the missing number.

21. In Prob. 3, suppose we want to assert, with 99 percent confidence, that the average salary is greater than v_1. What is the appropriate value of v_1? What would the value of v_2 be if we wanted to assert, with 99 percent confidence, that the average salary was less than v_2?

22. In Prob. 2, find a number that is, with 95 percent confidence, greater than the average time it takes California customers to receive their orders.

23. To convince a potential buyer of the worth of her company, an executive has ordered a survey of the daily cash receipts of the business. A sample of 14 days revealed the following values (in units of $100):

33, 12, 48, 40, 26, 17, 29, 38, 34, 41, 25, 51, 49, 34

If the executive wants to present these data in the most favorable way, should she present a confidence interval estimate or a one-sided confidence bound? If one-sided, should it be an upper or a lower bound? If you were the executive, how would you complete the following? "I am 95 percent confident that …"

24. To calm the concerns of a group of citizens worried about air pollution in their neighborhood, a government inspector has obtained sample data relating to carbon monoxide concentrations. The data, in parts per million, are as follows:

101.4, 103.3, 101.6, 111.6, 98.4, 95.0, 93.6

If these numbers appear reasonably low to the inspector, how should he, when speaking "with 99 percent confidence," present the results to the group?

8.7 INTERVAL ESTIMATORS OF A POPULATION PROPORTION

Suppose that we desire an interval estimator of p, the proportion of individuals in a large population who have a certain characteristic. Suppose further that a random sample of size n is chosen, and it is determined that X of the individuals in the sample have the characteristic. If we let $\hat{p} = X/n$ denote the proportion of the sample having the characteristic, then as previously noted in Sec. 8.3, the expected value and standard deviation of $\hat{p}$ are

$$E[\hat{p}] = p$$

$$\mathrm{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

When n is large enough that both np and $n(1-p)$ are greater than 5, we can use the normal approximation to the binomial distribution to assert that an approximate $100(1-\alpha)$ percent confidence interval estimator of p is given by

$$\hat{p} \pm z_{\alpha/2}\mathrm{SD}(\hat{p})$$

Although the standard deviation of $\hat{p}$ is not known, since it involves the unknown proportion p, we can estimate it by replacing p by its estimator $\hat{p}$ in the expression for $\mathrm{SD}(\hat{p})$. That is, we can estimate $\mathrm{SD}(\hat{p})$ by $\sqrt{\hat{p}(1-\hat{p})/n}$. This gives rise to the following.

An approximate $100(1-\alpha)$ percent confidence interval estimator of p is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p}$ is the proportion of members of the sample of size n who have the characteristic of interest.

■ Example 8.15

Out of a random sample of 100 students at a university, 82 stated that they were nonsmokers. Based on this, construct a 99 percent confidence interval estimate of p, the proportion of all the students at the university who are nonsmokers.

Solution

Since $100(1-\alpha) = 0.99$ when $\alpha = 0.01$, we need the value of $z_{\alpha/2} = z_{0.005}$, which from Table D.2 is equal to 2.576. The 99 percent confidence interval estimate of p is thus

$$0.82 \pm 2.576\sqrt{\frac{0.82(1-0.82)}{100}}$$

or

$$0.82 \pm 0.099$$

That is, we can assert with 99 percent confidence that the true percentage of nonsmokers is between 72.1 and 91.9 percent. ■

■ Example 8.16

On December 24, 1991, *The New York Times* reported that a poll indicated that 46 percent of the population was in favor of the way that President Bush was handling the economy, with a margin of error of ±3 percent. What does this mean? Can we infer how many people were questioned?

Solution

It has become common practice for the news media to present 95 percent confidence intervals. That is, unless it is specifically mentioned otherwise, it is almost always the case that the interval quoted represents a 95 percent confidence interval. Since $Z_{0.025} = 1.96$, a 95 percent confidence interval for p is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where n is the sample size. Since $\hat{p}$, the proportion of those in the random sample who are in favor of the President's handling of the economy, is equal to 0.46, it follows that the 95 percent confidence interval estimate of p, the proportion of the population in favor, is

$$0.46 \pm 1.96\sqrt{\frac{(0.46)(0.54)}{n}}$$

Since the margin of error is ±3 percent, it follows that

$$1.96\sqrt{\frac{(0.46)(0.54)}{n}} = 0.03$$

Squaring both sides of this equation shows that

$$(1.96)^2 \frac{(0.46)(0.54)}{n} = (0.03)^2$$

or

$$n = \frac{(1.96)^2(0.46)(0.54)}{(0.03)^2} = 1060.3$$

That is, approximately 1060 people were sampled, and 46 percent were in favor of President Bush's handling of the economy. ■

8.7.1 Length of the Confidence Interval

Since the $100(1-\alpha)$ percent confidence interval for p goes from

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{to} \quad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

it follows that the length of the interval is as follows.

Statistics in Perspective

Case Study

The Aid to Families with Dependent Children (AFDC) program recognizes that errors are inevitable, and so not every family that it funds actually meets the eligibility requirements. However, California holds its counties responsible for overseeing the eligibility requirements and has set a maximum error rate of 4 percent.That is, if over 4 percent of the funded cases in a county are found to be ineligible, then a financial penalty is placed upon the county, with the amount of the penalty determined by the error percentage. Since the state does not have the resources to check every case for eligibility, it uses random sampling to estimate the error percentages.

In 1981, a random sample of 152 cases was chosen in Alameda County, California, and 9 were found to be ineligible. Based on this estimated percentage of $100 \times 9/152 = 5.9$ percent, a penalty of $949,597 was imposed by the state on the county. The county appealed to the courts, arguing that 9 errors in 152 trials were not sufficient evidence to prove that its error percentage exceeded 4 percent. With help from statistical experts, the court decided that it was unfair to take the point estimate of 5.9 percent as the true error percentage of the county. The court decided it would be fairer to use the lower end of the 95 percent confidence interval estimate. Since the 95 percent confidence interval estimate of the proportion of all funded cases that are ineligible is

$$0.059 \pm 1.96\sqrt{\frac{0.059(1-0.059)}{152}} = 0.059 \pm 0.037$$

Statistics in Perspective (continued)

it follows that the lower end of this interval is $0.059 - 0.037 = 0.022$. Since this lower confidence limit is less than the acceptable value of 0.04, the court overturned the state's decision and ruled that no penalties were owed.

The length of a $100(1-\alpha)$ percent confidence interval is

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

where $\hat{p}$ is the proportion of the sample having the characteristic.

Since it can be shown that the product $\hat{p}(1-\hat{p})$ is always less than or equal to $1/4$, it follows from the preceding expression that an upper bound on the length of the confidence interval is given by $2z_{\alpha/2}\sqrt{1/(4n)}$, which is equivalent to the statement

$$\text{Length of } 100(1-\alpha) \text{ percent confidence interval} \leq \frac{z_{\alpha/2}}{\sqrt{n}}$$

The preceding bound can be used to determine the appropriate sample size needed to obtain a confidence interval whose length is less than a specified value. For instance, suppose that we want to determine a sufficient sample size so that the length of the resulting $100(1-\alpha)$ percent confidence interval is less than some fixed value b. In this case, upon using the preceding inequality, we can conclude that any sample size n for which

$$\frac{z_{\alpha/2}}{\sqrt{n}} < b$$

will suffice. That is, n must be chosen so that

$$\sqrt{n} > \frac{z_{\alpha/2}}{b}$$

Upon squaring both sides, we see that n must be such that

$$n > \left(\frac{z_{\alpha/2}}{b}\right)^2$$

■ Example 8.17

How large a sample is needed to ensure that the length of the 90 percent confidence interval estimate of p is less than 0.01?

Solution

To guarantee that the length of the 90 percent confidence interval estimator is less than 0.01, we need to choose n so that

$$n > \left(\frac{z_{0.05}}{0.01}\right)^2$$

Since $z_{0.05} = 1.645$, this gives

$$n > (164.5)^2 = 27,062.25$$

That is, the sample size needs to be at least 27,063 to ensure that the length of the 90 percent confidence interval will be less than 0.01.

If we let L denote the length of the confidence interval for p,

$$\begin{array}{ccc} \leftarrow & L & \rightarrow \\ \hline \hat{p}-\frac{L}{2} & \hat{p} & \hat{p}+\frac{L}{2} \end{array}$$

then since this interval is centered at $\hat{p}$, it follows that $\hat{p}$ is within $L/2$ of any point in the interval. Therefore, if p lies in the interval, then the distance from $\hat{p}$ to p is at most $L/2$. In Example 8.17 we can thus assert, with 90 percent confidence, that for a sample size as large as 27,063 the observed sample proportion will be within 0.005 of the true population proportion. ■

8.7.2 Lower and Upper Confidence Bounds

Lower and upper confidence bounds for p are easily derived and are given as follows.

A $100(1-\alpha)$ percent lower confidence bound for p is given by

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

That is, with $100(1-\alpha)$ percent confidence, the proportion of the population that has the characteristic is greater than this value.

A $100(1-\alpha)$ percent upper confidence bound for p is given by

$$\hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

That is, with $100(1-\alpha)$ percent confidence, the proportion of the population that has the characteristic is less than this value.

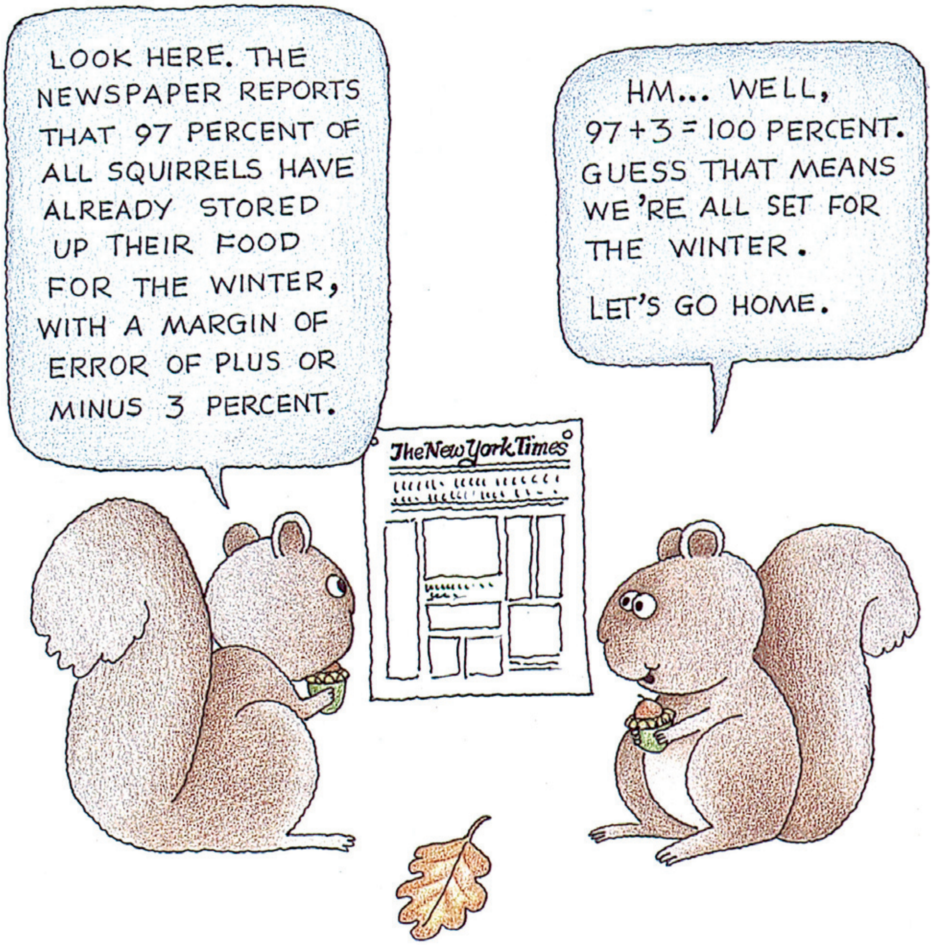
■ Example 8.18

A random sample of 125 individuals working in a large city indicated that 42 are dissatisfied with their working conditions. Construct a 95 percent lower confidence bound on the percentage of all workers in that city who are dissatisfied with their working conditions.

Solution

Since $z_{0.05} = 1.645$ and $42/125 = 0.336$, the 95 percent lower bound is given by

$$0.336 - 1.645\sqrt{\frac{0.336(0.664)}{125}} = 0.2665$$


LOOK HERE. THE NEWSPAPER REPORTS THAT 97 PERCENT OF ALL SQUIRRELS HAVE ALREADY STORED UP THEIR FOOD FOR THE WINTER, WITH A MARGIN OF ERROR OF PLUS OR MINUS 3 PERCENT.
HM... WELL, 97+3 = 100 PERCENT. GUESS THAT MEANS WE'RE ALL SET FOR THE WINTER.
LET'S GO HOME.
The New York Times

That is, we can be 95 percent certain that over 26.6 percent of all workers are dissatisfied with their working conditions. ■

PROBLEMS

1. A random sample of 500 California voters indicated that 302 are in favor of the death penalty. Construct a 99 percent confidence interval estimate of the proportion of all California voters in favor of the death penalty.
2. It is felt that first-time heart attack victims are particularly vulnerable to additional heart attacks during the year following the first attack. To estimate the proportion of victims who suffer an additional attack within 1 year, a random sample of 300 recent heart attack patients was tracked for 1 year.
   (a) If 46 of them suffered an attack within this year, give a 95 percent confidence interval estimate of the desired proportion.
   (b) Repeat part (a) assuming 92 suffered an attack within the year.
3. To estimate p, the proportion of all newborn babies who are male, the gender of 10,000 newborn babies was noted. If 5106 were male, determine a
   (a) 90 percent
   (b) 99 percent
   confidence interval estimate of p.
4. A poll of 1200 voters in 1980 gave Ronald Reagan 57 percent of the vote. Construct a 99 percent confidence interval estimate of the proportion of the population who favored Reagan at the time of the poll.
5. A random sample of 100 Los Angeles residents indicated that 64 were in favor of strict gun control legislation. Determine a 95 percent confidence interval estimate of the proportion of all Los Angeles residents who favor gun control.
6. A random sample of 100 recent recipients of Ph.D.s in science indicated that 42 were optimistic about their future possibilities in science. Find a
   (a) 90 percent
   (b) 99 percent
   confidence interval estimate of the proportion of all recent recipients of Ph.D.s in science who are optimistic.
7. In Prob. 1 of Sec. 8.3, find a 95 percent confidence interval estimate of the proportion of North Americans who believed the Communist party would have won a free election in the Soviet Union in 1985.
8. Using the data of Prob. 4 of Sec. 8.3, find a 90 percent confidence interval estimate of the probability of winning at solitaire.

9. A wine importer has the opportunity to purchase a large consignment of 1947 Chateau Lafite Rothschild wine. Because of the wine's age, some of the bottles may have turned to vinegar. However, the only way to determine whether a bottle is still good is to open it and drink some. As a result, the importer has arranged with the seller to randomly select and open 20 bottles. Suppose 3 of these bottles are spoiled. Construct a 95 percent confidence interval estimate of the proportion of the entire consignment that is spoiled.
10. A sample of 100 cups of coffee from a coffee machine is collected, and the amount of coffee in each cup is measured. Suppose that 9 cups contain less than the amount of coffee specified on the machine. Construct a 90 percent confidence interval estimate of the proportion of all cups dispensed that give less than the specified amount of coffee.
11. A random sample of 400 librarians included 335 women. Give a 95 percent confidence interval estimate of the proportion of all librarians who are women.
12. A random sample of 300 authors included 117 men. Give a 95 percent confidence interval estimate of the proportion of all authors who are men.
13. A random sample of 9 states (West Virginia, New York, Idaho, Texas, New Mexico, Indiana, Utah, Maryland, and Maine) indicated that in 2 of these states the 1990 per capita income exceeded $20,000. Construct a 90 percent confidence interval estimate of the proportion of all states that had a 1990 per capita income in excess of $20,000.
14. A random sample of 1000 psychologists included 457 men. Give a 95 percent confidence interval estimate of the proportion of all psychologists who are men.
15. A random sample of 500 accountants included 42 African Americans, 18 Hispanic Americans, and 246 women. Construct a 95 percent confidence interval estimate of the proportion of all accountants who are
    (a) African American
    (b) Hispanic American
    (c) Female
16. In a poll conducted on January 22, 2004, out of a random sample of 600 people, 450 stated they were in favor of the war against Iraq. Construct a
    (a) 90 percent
    (b) 95 percent
    (c) 99 percent

    confidence interval estimate of p, the proportion of the population in favor of the war at the time.

17. The poll mentioned in Prob. 16 was quoted in the January 28, 2004, *San Francisco Chronicle*, where it was stated that "75 percent of the population are in favor, with a margin of error of plus or minus 4 percentage points."
    (a) Explain why the *Chronicle* should have stated that the margin of error is plus or minus 3.46 percentage points.
    (b) Explain how the *Chronicle* erred to come up with the value of ±4 percent.
18. A recent newspaper poll indicated that candidate A is favored over candidate B by a 53-to-47 percentage, with a margin of error of ±4 percent. The newspaper then stated that since the 6-point gap is larger than the margin of error, its readers can be certain that candidate A is the current choice. Is this reasoning correct?
19. A market research firm is interested in determining the proportion of households that are watching a particular sporting event. To accomplish this task, it plans on using a telephone poll of randomly chosen households.
    (a) How large a sample is needed if the company wants to be 90 percent certain that its estimate is correct to within ±0.02?
    (b) Suppose there is a sample whose size is the answer in part (a). If 23 percent of the sample were watching the sporting event, do you expect that the 90 percent confidence interval will be exactly of length 0.02, larger than 0.02, or smaller than 0.02?
    (c) Construct the 90 percent confidence interval for part (b).
20. What is the smallest number of death certificates we must randomly sample to estimate the proportion of the U.S. population that dies of cancer, if we want the estimate to be correct to within 0.01 with 95 percent confidence?
21. Suppose in Prob. 20 that it is known that roughly 20 percent of all deaths are due to cancer. Using this information, determine approximately how many death certificates will have to be sampled to meet the requirements of Prob. 20.
22. Use the data of Prob. 14 to obtain a 95 percent lower confidence bound for the proportion of all psychologists who are men.
23. Use the data of Prob. 11 to obtain a 95 percent upper confidence bound for the proportion of all librarians who are women.
24. A manufacturer is planning on putting out an advertisement claiming that over x percent of the users of his product are satisfied with it. To determine x, a random sample of 500 users was questioned. If 92 percent of these people indicated satisfaction and the manufacturer wants to be 95 percent confident about the validity of the advertisement, what value of x should be used in the advertisement? What value

should be used if the manufacturer was willing to be only 90 percent confident about the accuracy of the advertisement?

25. Use the data of Prob. 15 to obtain a

    (a) 90 percent lower confidence bound

    (b) 90 percent upper confidence bound

    for the proportion of all accountants who are either African American or Hispanic American.

26. In Prob. 16 construct a

    (a) 95 percent upper confidence bound

    (b) 95 percent lower confidence bound

    for p, the proportion of the population in favor of the war at the time of the poll.

27. Suppose in Prob. 9 that the importer has decided that purchase of the consignment will be profitable if less than 20 percent of the bottles is spoiled. From the data of this problem, should the importer be

    (a) 95 percent certain

    (b) 99 percent certain

    that the purchase will be profitable?

28. Refer to the data in Prob. 5. Fill in the missing numbers for these statements:

    (a) With 95 percent confidence, more than _______ percent of all Los Angeles residents favor gun control.

    (b) With 95 percent confidence, less than _______ percent of all Los Angeles residents favor gun control.

KEY TERMS

Estimator: A statistic used to approximate a population parameter. Sometimes called a *point estimator*.

Estimate: The observed value of the estimator.

Unbiased estimator: An estimator whose expected value is equal to the parameter that it is trying to estimate.

Standard error of an (unbiased) estimator: The standard deviation of the estimator. It is an indication of how close we can expect the estimator to be to the parameter.

Confidence interval estimator: An interval whose endpoints are determined by the data. The parameter will lie within this interval with a certain degree of confidence. This interval is usually centered at the point estimator of the parameter.

$100(1 - \alpha)$ percent level of confidence: The long-term proportion of time that the parameter will lie within the interval. Equivalently, before the data are observed, the interval estimator will contain the parameter with probability $1 - \alpha$; after the data are observed, the resultant interval estimate contains the parameter with $100(1 - \alpha)$ percent *confidence*.

Lower confidence bound: A number, whose value is determined by the data, which is less than a certain parameter with a given degree of confidence.

Upper confidence bound: A number, whose value is determined by the data, which is greater than a certain parameter with a given degree of confidence.

***t* Random variable**: If $X_1, \ldots, X_n$ are a sample from a normal population having mean μ, then the random variable

$$\sqrt{n}\frac{\overline{X} - \mu}{S}$$

is said to be a t random variable with $n - 1$ degrees of freedom, where $\overline{X}$ and S are, respectively, the sample mean and sample standard deviation.

SUMMARY

The sample mean $\overline{X}$ is an unbiased estimator of the population mean μ. Its standard deviation, sometimes referred to as the *standard error* of $\overline{X}$ as an estimator of μ, is given by

$$\text{SD}(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

where σ is the population standard deviation.

The statistic $\hat{p}$, equal to the proportion of a random sample having a given characteristic, is the estimate of p, the proportion of the entire population with the characteristic. The standard error of the estimate is

$$\text{SD}(\hat{p}) = \sqrt{\frac{p(1 - p)}{n}}$$

where n is the sample size. The standard error can be estimated by

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

The sample variance S^2 is the estimator of the population variance σ^2. Correspondingly, the sample standard deviation S is used to estimate the population standard deviation σ.

If $X_1, \ldots, X_n$ are a sample from a normal population having a known standard deviation σ,

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

is a $100(1-\alpha)$ percent confidence interval estimator of the population mean μ. The length of this interval, namely,

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

will be less than or equal to b when the sample size n is such that

$$n \geq \left(\frac{2z_{\alpha/2}\sigma}{b}\right)^2$$

A $100(1-\alpha)$ lower confidence bound for μ is given by

$$\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

That is, we can assert with $100(1-\alpha)$ percent confidence that

$$\mu > \overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

A $100(1-\alpha)$ upper confidence bound for μ is

$$\overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

That is, we can assert with $100(1-\alpha)$ percent confidence that

$$\mu < \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

If $X_1, \ldots, X_n$ *are a sample from a normal population whose standard deviation is unknown*, a $100(1-\alpha)$ percent confidence interval estimator of μ is

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

In the preceding, $t_{n-1,\alpha/2}$ is such that

$$P\left\{T_{n-1} > t_{n-1,\alpha/2}\right\} = \frac{\alpha}{2}$$

when T_{n-1} is a t random variable with $n-1$ degrees of freedom.

The $100(1-\alpha)$ percent lower and upper confidence bounds for μ are, respectively, given by

$$\overline{X} - t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

and

$$\overline{X} + t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

To obtain a confidence interval estimate of p, the proportion of a large population with a specific characteristic, take a random sample of size n. If $\hat{p}$ is the proportion of the random sample that has the characteristic, then an approximate $100(1-\alpha)$ percent confidence interval estimator of p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The length of this interval always satisfies

$$\text{Length of confidence interval} \leq \frac{z_{\alpha/2}}{\sqrt{n}}$$

The distance from the center to the endpoints of the 95 percent confidence interval estimator, that is, $1.96\sqrt{\hat{p}(1-\hat{p})/n}$, is commonly referred to as the *margin of error*. For instance, suppose a newspaper states that a new poll indicates that 64 percent of the population consider themselves to be conservationists, with a margin of error of ±3 percent. By this, the newspaper means that the results of the poll yield that the 95 percent confidence interval estimate of the proportion of the population who consider themselves to be conservationists is 0.64 ± 0.03.

REVIEW PROBLEMS

1. Which case would yield a more precise estimator of μ?
   - (a) A sample of size n from a population having mean 2μ and variance σ^2
   - (b) A sample of size $2n$ from a population having mean μ and standard deviation σ
2. The weights of ball bearings are normally distributed with standard deviation 0.5 millimeters.
   - (a) How large a sample is needed if you want to be 95 percent certain that your estimate of the mean weight of a ball bearing is correct to within ±0.1 millimeters?
   - (b) Repeat (a) if you want the estimate to be correct to within ±0.01 millimeters.
   - (c) If a sample of size 8 yields the values

     4.1, 4.6, 3.9, 3.3, 4.0, 3.5, 3.9, 4.2

     give a 95 percent confidence interval estimate for the mean weight.

3. A random sample of 50 people from a certain population was asked to keep a record of the amount of time spent watching television in a specified week. If the sample mean of the resulting data was 24.4 hours and the sample standard deviation was 7.4 hours, give a 95 percent confidence interval estimate for the average time spent watching television by all members of the population that week.
4. Use the first 30 data values in App. A to give a 90 percent confidence interval estimate of the average blood cholesterol level of all the students on the list. Now break up the 30 data values into two groups—one for the females and one for the males. Use the data for each gender separately to obtain 90 percent confidence interval estimates for the mean cholesterol level of the women and of the men. How much confidence would you put in the assertion that the average levels for both the men and the women lie within their respective 90 percent confidence intervals?
5. A standardized test is given annually to all sixth-grade students in the state of Washington. To find out the average score of students in her district, a school supervisor selects a random sample of 100 students. If the sample mean of these students' scores is 320 and the sample standard deviation is 16, give a 95 percent confidence interval estimate of the average score of students in that supervisor's district.
6. An airline is interested in determining the proportion of its customers who are flying for reasons of business. If the airline wants to be 90 percent certain that its estimate will be correct to within 2 percent, how large a random sample should it select?
7. The following data represent the number of drinks sold from a vending machine on a sample of 20 days:

   56, 44, 53, 40, 65, 39, 36, 41, 47, 55, 51, 50, 72, 45, 69, 38, 40, 51, 47, 53

   (a) Determine a 95 percent confidence interval estimate of the mean number of drinks sold daily.
   (b) Repeat part (a) for a 90 percent confidence interval.
8. It is thought that the deepest part of sleep, which is also thought to be the time during which dreams most frequently occur, is characterized by rapid eye movement (REM) of the sleeper. The successive lengths of seven REM intervals of a sleep volunteer were determined at a sleep clinic. The following times in minutes resulted:

   37, 42, 51, 39, 44, 48, 29

   Give a 99 percent confidence interval estimate for the mean length of a REM interval of the volunteer.
9. A large corporation is analyzing its present health care policy. It is particularly interested in its average cost for delivering a baby.

Suppose that a random sample of 24 claims yields that the sample mean of the delivery costs is \$1840 and the sample standard deviation is \$740. Construct a 95 percent confidence interval estimate of the corporation's present mean cost per delivery.

10. A court-ordered survey yielded the result that out of a randomly chosen sample of 300 farm workers, 144 were in favor of unionizing. Construct a 90 percent confidence interval estimate of the proportion of all farm workers who wanted to be unionized.

11. A sample of nine fastballs thrown by a certain pitcher were measured at speeds of

$$94, 87, 80, 91, 85, 102, 85, 80, 93$$

miles per hour.

(a) What is the point estimate of the mean speed of this pitcher's fastball?

(b) Construct a 95 percent confidence interval estimate of the mean speed.

12. A sample of size 9 yields a sample mean of 35. Construct a 95 percent confidence interval estimate of the population mean if the population standard deviation is known to equal

(a) 3

(b) 6

(c) 12

13. Repeat Prob. 12 for a sample size of 36.

14. The following are scores on IQ tests of a random sample of 18 students at a large eastern university:

$$130, 122, 119, 142, 136, 127, 120, 152, 141,$$
$$132, 127, 118, 150, 141, 133, 137, 129, 142$$

(a) Construct a 90 percent confidence interval estimate of the average IQ score of all students at the university.

(b) Construct a 95 percent confidence interval estimate.

(c) Construct a 99 percent confidence interval estimate.

15. To comply with federal regulations, the state director of education needs to estimate the proportion of all secondary school teachers who are female. If there are 518 females in a random sample of 1000 teachers, construct a 95 percent confidence interval estimate.

16. In Prob. 15, suppose the director had wanted a 99 percent confidence interval estimate whose length was guaranteed to be at most 0.03. How large a sample would have been necessary?

17. The Census Bureau, to determine the national unemployment rate, uses a random sample of size 50,000. What is the largest possible margin of error?

18. A researcher wants to learn the proportion of the public that favors a certain candidate for office. If a random sample of size 1600 is chosen, what is the largest possible margin of error?

19. A problem of interest in baseball is whether a sacrifice bunt is a good strategy when there is a player on first base and there are no outs. Assuming that the bunter will be out but will be successful in advancing the runner on base, we could compare the probability of scoring a run with a player on first base and no outs to the probability of scoring a run with a player on second base and one out. The following data resulted from a study of randomly chosen major league baseball games played in 1959 and 1960.

Base occupied	Number of outs	Proportion of cases in which no runs are scored	Total number of cases
First	0	0.604	1728
Second	1	0.610	657

(a) Give a 95 percent confidence interval estimate for the probability of scoring at least one run when there is a player on first base and there are no outs.

(b) Give a 95 percent confidence interval estimate for the probability of scoring at least one run when there is a player on second base and one out.

20. Use the data of Prob. 15 to construct a

(a) 90 percent

(b) 95 percent

(c) 99 percent

upper confidence bound for the proportion of all secondary school teachers who are female.

21. Repeat Prob. 20, this time constructing lower confidence bounds. If you were an advocate of greater hiring of female teachers, would you tend to quote an upper or a lower confidence bound?

22. Suppose that a random sample of nine recently sold houses in a certain neighborhood resulted in a sample mean price of \$222,000, with a sample standard deviation of \$12,000. Give a 95 percent upper confidence bound for the mean price of all recently sold houses in this neighborhood.