

To be assured of finding the desired level of confidence, always round up. Thus, we are 90% confident that a sample of  $n = 98$  observations would produce an estimate of the mean amount of cleaning fluid in a 12-ounce bottle to within 0.05 ounce. Being able to know the accuracy of your estimate is one of the significant benefits of inferential statistics. In this case, if 98 bottles are measured, we will be 90% confident that the resulting sample mean is within five one-hundredths of an ounce of the true mean. That's close.

### Determining the Sample Size: $\sigma$ Unknown

In the previous discussion of determining the sample size necessary to estimate a population mean with a desired accuracy,  $\sigma$  was assumed to be known. This assumption is usually unreasonable in most problem-solving environments.

The most obvious method for obtaining an estimate of  $\sigma$  is to take a small sample and use the sample standard deviation as an estimate of the population standard deviation. Replacing  $\sigma$  with  $s$  in the sample size determination relationship will provide an initial estimate of the required sample size. Another alternative is to use the value of the sample standard deviation obtained in a previous study, sometimes called a **pilot study**. Keep in mind that in order to construct a confidence interval for a population mean when the population standard deviation is unknown, the distribution of the population is assumed to be normal.

#### Example 9.2.5

#### Calculating the Sample Size Needed for 99% Confidence of the Mean Time Required to Replace a Jet Engine

An airline's maintenance manager desires to estimate the average time (in hours) required to replace a jet engine in a Boeing 767. How large a sample would be necessary if the manager wishes to be 99% confident of estimating the population mean to within one-quarter of an hour ( $E = 0.25$ )? Assume a preliminary sample of size  $n = 30$  has a mean replacement time of 16.7 hours with a standard deviation of 4.3 hours.

#### SOLUTION

Using the results from the initial sample,

$$n = \left( \frac{z_{\alpha/2} s}{E} \right)^2 = \left( \frac{2.575 \cdot 4.3}{0.25} \right)^2 = 1961.6041$$

$$n = 1962 \text{ (Always round up to assure required confidence.)}$$

Notice that while the sample data values are being collected they can be used to improve the estimate of the population standard deviation. For example, suppose the sample standard deviation after sampling the first 1000 observations was 4.1. Using this estimate of  $s$  instead of 4.3 results in a sample size of 1784 compared to the original specification of 1962. The notion of modifying the sample size estimate as additional data are observed can be applied at regular intervals during the sampling process until the estimate of the standard deviation stabilizes.

#### Six Degrees of Separation: A Law of Small Worlds

What is the number of people a randomly chosen person in Omaha, Nebraska needs to contact before she can find a connection with a randomly chosen housewife in New England? How many intervening people do you think separates you from the President of the United States? Unsuspecting readers might guess very large numbers but the actual numbers are quite small. The answer to both of these questions may very well be less than 6! Psychologists have done ingenious experiments and have actually calculated this degree of separation, on average, to be six. What is amazing about this degree of separation is that it is equally true for the President of the United States and a sweet vendor in Bangladesh.

*continued on next page...*

## 9.2 Exercises

### Basic Concepts

1. Why is the assumption that the population standard deviation is known when estimating the population mean not very realistic?
2. What effect does knowing the standard deviation of the population have on the construction of the confidence interval?

3. What is the Student's  $t$ -distribution?
4. What are the conditions in which the  $t$ -distribution is used in interval estimation of the population mean?
5. What is the parameter of the  $t$ -distribution? How is it calculated?
6. What is the value of having a confidence interval with a small width?
7. Can a confidence interval be constructed with a width of your choice? Explain.
8. What are the three components that affect the width of the confidence interval for the population mean? Describe how changes in these three components affect the width of the confidence interval.
9. What is the margin of error? What is the connection between the expression for the margin of error and the equation to determine the sample size?
10. What is the rounding rule regarding the determination of the sample size?
11. What is the difference between the method of determining the sample size when  $\sigma$  is known versus when  $\sigma$  is unknown?
12. What is a pilot study?
13. Note that  $E$  and  $s$  can be viewed as measures of variation. Compare and contrast the meanings of  $E$  and  $s$  in layman's terms.

*continued...*

The degree of separation for the number of clicks that you will need to make to get to a website that interests you, as well as the analysis of terrorist networks, turn out to be of similar nature. The new science of networks can shed useful light and help to derive general laws applicable to many of these types of questions.

### Exercises

14. Find the  $t$ -value such that 0.025 of the area under the curve is to the right of the  $t$ -value. Assume the degrees of freedom equal 13.
15. Find the  $t$ -value such that 0.01 of the area under the curve is to the right of the  $t$ -value. Assume the degrees of freedom equal 21.
16. Find  $t_{\alpha/2, df}$  for the following combinations of  $\alpha$  and  $n$ .
  - a.  $\alpha = 0.05, n = 15$
  - b.  $\alpha = 0.01, n = 20$
  - c.  $\alpha = 0.10, n = 8$
17. Find  $t_{\alpha/2, df}$  for the following combinations of  $\alpha$  and  $n$ .
  - a.  $\alpha = 0.05, n = 12$
  - b.  $\alpha = 0.01, n = 18$
  - c.  $\alpha = 0.10, n = 22$
18. A random sample, consisting of the values listed below, was taken from a normally distributed population. Assuming the standard deviation of the population is unknown, construct a 99% confidence interval for the population mean.

27.4	26.5	25.7	31.4
28.2	21.9	16.3	22.7
18.8	34.4	29.2	20.5

19. Construct an 80% confidence interval for the mean of a normal population assuming that the values listed below comprise a random sample taken from the population. The population standard deviation is unknown.

83.9	87.4	65.2	86.0	73.1
80.3	92.7	87.5	69.3	77.5
91.9	71.1	79.1	72.4	88.2

20. An FDA representative randomly selects 8 packages of ground chuck from a grocery store and measures the fat content (as a percent) of each package. The resulting measurements are given below.

Fat Contents			
13%	12%	14%	17%
15%	16%	18%	15%

- Calculate the sample mean and the sample standard deviation of the fat contents.
  - Construct a 90% confidence interval for the true mean fat content of all the packages of ground beef.
  - What assumption did you make about the fat content in constructing your interval?
21. A hospital would like to determine the mean length of stay for its patients having abdominal surgery. A sample of 15 patients revealed a sample mean of 6.4 days and a sample standard deviation of 1.4 days.
- Find a 95% confidence interval for the mean length of stay for patients with abdominal surgery.
  - Interpret this interval and state any assumptions that were made in the construction of the interval.
22. An independent group of food service personnel conducted a survey on tipping practices in a large metropolitan area. They collected information on the percentage of the bill left as a tip for 25 randomly selected bills. The average tip was 12.3% of the bill with a standard deviation of 2.7%.
- Construct an interval to estimate the true average tip (as a percent of the bill) with 99% confidence.
  - Interpret the interval, and state any assumptions that were made in the construction of the interval.
23. A travel agent is interested in the average price of a hotel room during the summer in a resort community. The agent randomly selects 15 hotels from the community and determines the price of a regular room with a king size bed. The average price of the room for the sample was \$115 with a standard deviation of \$30.
- Construct an interval to estimate the true average price of a regular room with a king size bed in the resort community with 90% confidence.
  - Interpret the interval, and state any assumptions that were made in the construction of the interval.
24. In 2010 the median home price in all regions of the United States was \$221,800. It is commonly thought that better schools are found in wealthier areas. In *Forbes* magazine's list of the "Best Schools for your Real Estate Buck," the top 10 cities in America were identified where your housing dollar will go the furthest in getting your children a great education. 17,589 towns and cities were analyzed using results from the most recent National Assessment for Educational Progress data, and the top 10 school districts were identified. The list counteracted the idea that more money equals better schools, as Falmouth, Maine topped the list beating out high-dollar school districts like Manhattan Beach, California. The top 10 cities are given below, along with the median home price for each city.

Best Schools for Your Real Estate Buck		
Education Rank	City	Median Home Price (\$)
1	Falmouth, Maine	351,550
2	Mercer Island, Washington	708,740
3	Pella, Iowa	148,200

Best Schools for Your Real Estate Buck (cont.)		
Education Rank	City	Median Home Price (\$)
4	Barrington, Rhode Island	296,010
5	Bedford, New Hampshire	293,730
6	Manhattan Beach, California	1,278,980
7	Moraga, California	722,010
8	Parkland, Florida	426,390
9	St. Johns, Florida	181,700
10	Southlake, Texas	476,880

Source: Forbes magazine

- a. Construct a 90% confidence interval for the median home price of cities on the top 10 list.
  - b. Is the average median price for these cities higher than the median price for the U.S. as a whole?
  - c. What population assumption needs to be made here?
  - d. How would your solutions to **a.** and **b.** change if these were mean rather than median values?
25. A technician working for the Chase-National Food Additive Company would like to estimate the preserving ability of a new additive. This additive will be used for Auntie's brand preserves. Based on past tests, it is believed that the time to spoilage for this additive has a standard deviation of 6 days. To be 90% confident of the true mean time to spoilage, what sample size will be needed to estimate the mean time to spoilage with an accuracy of one day?
26. A computer software company would like to estimate how long it will take a beginner to become proficient at creating a graph using their new spreadsheet package. Past experience has indicated that the time required for a beginner to become proficient with a particular function of the new software product has an approximately normal distribution with a standard deviation of 15 minutes. Find the sample size necessary to estimate the true average time required for a beginner to become proficient at creating a graph with the new spreadsheet package to within 5 minutes with 95% confidence.
27. A hot-dog vendor is evaluating a downtown location by counting the number of people who walk past the prospective location on a particular day during lunch time (i.e. 11:00 AM to 2:00 PM). A preliminary study has indicated a standard deviation of about 30 people per lunch period. How many lunch periods will be needed to estimate the average number of people who walk past the prospective location during the lunch period to within 9 people with 90% confidence?

## 9.3 Estimating the Population Proportion

An attribute is a characteristic that members of a population either possess or do not possess. Attributes are almost always measured as the **proportion** of the population that possesses the characteristic.

Many decisions require a measure of a population attribute. Television and radio stations base their advertising charges on ratings reflecting the *percentage of television viewers who are watching a particular program*. A political analyst wants to know the *fraction of voters who favor a particular candidate*. A social researcher needs the *fraction of teachers who believe group learning is a beneficial instructional method*. An insurance company is interested in estimating the *fraction of their policies that will result in claims*. A quality control engineer requires the *percentage defective in a lot of goods*. A marketing researcher demands the