Confidence Builder

When a light-bulb company reports a mean life expectancy for its products, how confident can you be that their bulbs won't leave you in the dark?

Danny Jones • Kate Riley • John Knudson-Martin

Confidence Builder

Introduction

To obtain reliable information on populations, researchers depend on sampling techniques. The statistics generated through sampling typically are used to predict the parameters of a population. For example, political parties and government officials use the results of polls to help them gauge public opinion. Manufacturers rely on sampling to monitor the quality of their products. And economists use business statistics to study market trends.

Discussion

a. Describe the differences between a statistic and a parameter.

b. Describe some populations in your school and in your community.

c. What parameters might be of interest for each population described in Part **b**?

d.
  1. Describe some different ways in which samples could be taken from the populations in Part **b**.

  2. Which of these sampling methods generate simple random samples?

Activity 1

Imagine that your school board is planning to purchase some new student desks. Your class has been asked to describe the dimensions of a comfortable desk. Since these desks must accommodate a wide range of students, the class decides to investigate student heights first.

Some classmates suggest that the heights of students in the class could be used to estimate the mean height of all students in the school. Others argue that since the class is not a representative sample of the student population, the sample mean will not provide a good estimate of the population mean. How could you determine if the class mean is a reasonable approximation of the mean height of the entire student population?

Mathematics Note

The mean value for a population, or **population mean**, is denoted by the Greek letter μ (mu).

The **population standard deviation** is denoted by the Greek letter σ (sigma). It can be calculated using the following formula:

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}}$$

where the population has *N* members represented by $x_1, x_2, \ldots, x_N$.

The mean value for a sample, or **sample mean**, is denoted by $\bar{x}$ (read "x-bar"). The **sample standard deviation**, denoted by *s*, can be calculated as follows:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}}$$

where the sample has *n* members represented by $x_1, x_2, \ldots, x_n$.

Notice that the denominator used to calculate the sample standard deviation is slightly different from the denominator used to calculate the population standard deviation. When calculating sample standard deviation, the denominator $n - 1$ provides a better estimate of the population standard deviation.

For example, consider a population of the digits from 0 to 9. The population mean μ is 4.5. The population standard deviation σ is approximately 2.87. For the following sample of 5 digits taken from this population—4, 6, 8, 7, 9—the sample mean $\bar{x}$ is 6.8. The sample standard deviation s is approximately 1.9.

Exploration

In this exploration, you use sampling techniques to estimate the mean height of your class.

a. Select a simple random sample of students in your class. Measure and record the height (in centimeters) of each student in the sample.

b. Calculate the mean height ($\bar{x}$) and the standard deviation (*s*) of the sample data.

c. Measure and record the height (in centimeters) of all students in the class population.

d. Calculate the mean height (μ) and the standard deviation (σ) of the class data.

e. As shown in Figure **1**, a relative frequency histogram consists of bars of equal width whose heights indicate the relative frequencies of measurements in the corresponding intervals.

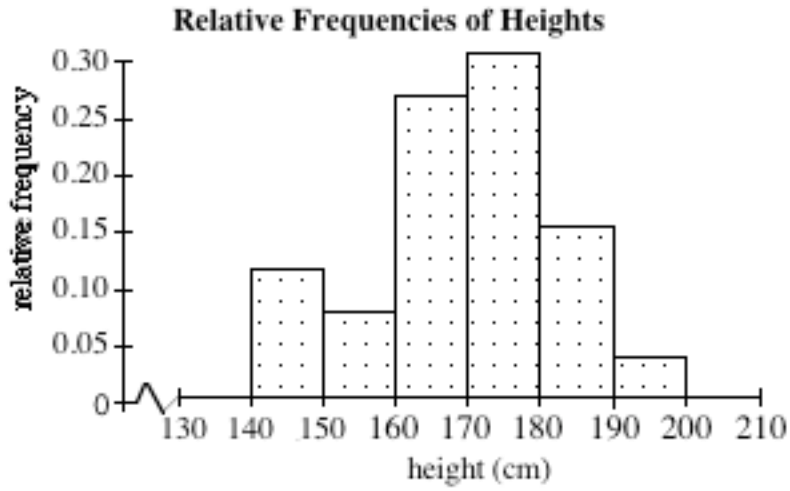


Figure 1: A relative frequency histogram

1. Create a relative frequency histogram of the class data using appropriate intervals for student heights.
2. Sketch the relative frequency polygon of the class data on the histogram from Step **1**.

Discussion

a.

1. How well did the sample mean approximate the actual mean height of the class?
2. What factors might have affected the accuracy of the sample mean as an estimate of the population mean?
3. Do you think that the heights of students in the class represent an unbiased sample of the heights of students in the entire school? Explain your response.

b. Describe the shape of the relative frequency polygon created in Part **e** of the exploration.

c. How would the shape of the relative frequency polygon change if the population consisted of professional basketball players?

Mathematics Note

A **normal distribution** is a continuous probability distribution. The graph of a normal distribution is symmetric about the mean and tapers to the left and right like a bell. The curve that describes the shape of the graph is the **normal curve**. As in all continuous probability distributions, the total area between the *x*-axis and a normal curve is 1.

Although all normal curves have the same general shape, the width of any particular curve depends on the standard deviation (σ) of the distribution that the curve models. For example, Figure **2** shows two normal curves that have the same mean but different standard deviations.

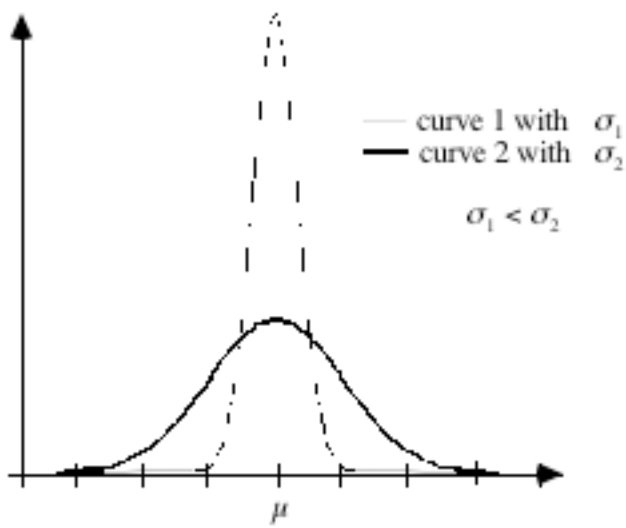


Figure 2: Two normal curves with the same mean

Since the population mean (μ) is located at the point where the curve's axis of symmetry intersects the *x*-axis, the position of the curve along the *x*-axis depends on the value of μ. Two normal curves with different means but the same standard deviation are shown in Figure **3**.

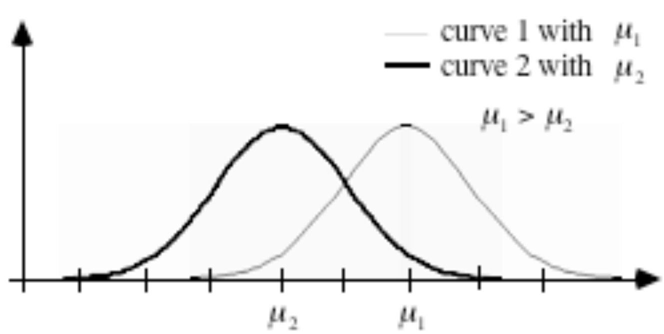


Figure 3: Two normal curves with the same standard deviation

d.

1. What differences do you observe between the two normal curves in Figure **2**?
2. What differences do you observe between the two normal curves in Figure **3**?

e. Describe how the value of σ affects the shape of a normal curve.

f. Describe how the value of μ affects the position of a normal curve.

Assignment

1.1

a. For a survey on the size of local households, record the number of people living in each of five different households in your community.

b. Calculate the mean and standard deviation of your data.

c. Describe any biases your sample may contain with respect to each of the following populations:

1. your community
2. the United States
3. the world.

1.2 Write the whole numbers from 1 to 30, one at a time, on 30 identical slips of paper. Place the slips of paper in a container.

a. What is the mean of this population?

b. Select a random sample of two numbers from the container. Find the mean of the sample.

c. Repeat Part **b** for a sample of 5 numbers and a sample of 20 numbers.

d. Compare the three sample means to the population mean. Describe any trends you observe.

Mathematics Note

The **law of large numbers** states that, for very large sample sizes, there is a high probability that the sample mean is close to the population mean.

For example, suppose that you want to estimate the mean number of hours spent on homework in a population of 150 students. If you plan to sample this population, a sample of 50 students is likely to provide a better estimate of the population mean than a sample of 5 students.

1.3 **a.** Consider the whole numbers from 1 to 99 as a population. Using appropriate technology, generate three random samples from this population: one sample of 10 numbers, one of 20 numbers, and one of 80 numbers. Find the mean of each sample.

b. Repeat Part **a** at least three more times.

c. Compare the three sample means for each sample size to the population mean, $\mu = 50$. In a paragraph, describe how the results of this experiment relate to the law of large numbers.

* * * * *

1.4 The following table shows the summer earnings of a population of students.

\$1872	\$1341	\$1792	\$1650	\$1422
\$1413	\$1900	\$2143	\$786	\$451
\$2432	\$0	\$243	\$1381	\$187
\$0	\$2443	\$1408	\$187	\$0
\$1228	\$1119	\$748	\$949	\$2011
\$896	\$1740	\$0	\$483	\$846
\$556	\$780	\$314	\$768	\$635

a. Calculate the mean and standard deviation of the data.

b. **1.** Select a random sample of five values from this population.

2. Calculate the mean and standard deviation of the sample.

c. **1.** Is your sample mean a good estimate of the population mean?

2. How could you select a sample that would provide a better estimate?

* * * * * * * * * *

Activity 2

How well do you think the mean height of your class estimates the mean height of your entire school? Your answer is likely to depend on the size of the school population. Since finding parameters for a large population is usually difficult (if not impossible), researchers use a variety of sampling techniques to obtain estimates.

In this activity, you examine how sample size can affect the distribution of sample means.

Exploration

A biologist has just received a shipment of 90 live fish for an upcoming research project. The hatchery claims that the mean length of fish in the shipment is 25 cm. The biologist, however, thinks that the fish look a little on the small side. Although it would be too much trouble to catch and measure every fish, you can help the biologist decide whether or not to believe the hatchery by sampling the population.

a.
   1. Using the template provided by your teacher, calculate the mean length and standard deviation of the fish population.
   2. Create a relative frequency polygon that represents the distribution of the data.

b. Cut the numbers representing fish lengths from the template and place them in a container. Mix the numbers thoroughly.

c. Select a simple random sample of 30 "fish" from your "pond." Record the length of one fish at a time, returning it to the pond before drawing the next. Be sure to mix the fish before each draw.

d. Calculate the mean and standard deviation of your sample.

e. Repeat Parts **c** and **d** nine more times.

f. Find the mean and standard deviation of your set of 10 sample means.

g. Create a relative frequency polygon that represents the distribution of the 10 sample means.

h. Collect the sample means of the entire class. Create a relative frequency polygon that represents the distribution of the class data.

Discussion

a. How did the mean and standard deviation of your 10 sample means in Part **f** of the exploration compare with those of your classmates?

b. How did the mean and standard deviation of the sample means calculated in Part **f** of the exploration compare to the mean and standard deviation of the original population?

c. Do you think that the mean length claimed by hatchery officials was a reasonable estimate of the population mean?

d. How did the relative frequency polygon you created in Part **g** of the exploration compare with those of your classmates?

e. Describe the relative frequency polygon created in Part **h** of the exploration using the class statistics.

Mathematics Note

The **central limit theorem** states that even if the population from which samples are taken is not normally distributed, the distribution of the means of all possible samples of the same size will be approximately normal. In other words, if you collect many samples of size n and create a relative frequency histogram and polygon of the sample means, the graph will tend to assume the bell shape of a normal curve. Figure **4** shows an example of such a distribution.

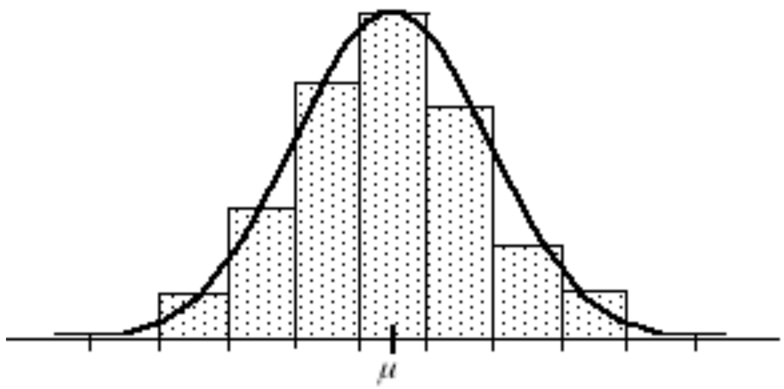


Figure 4: Normal distribution of sample means

This approximation becomes more accurate as the sample size n increases. Statisticians generally agree that for $n \geq 30$, the distribution of sample means can be modeled reasonably well by a normal curve. This requirement is not necessary if the population itself is normally distributed.

The mean of the distribution of all sample means equals μ, the population mean. The **standard deviation of all sample means**, denoted $\sigma_{\bar{x}}$, can be calculated using the following formula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation of the population and n is the sample size.

For example, consider a population in which some characteristic is not normally distributed, with $\mu = 35$ and $\sigma = 5$. The standard deviation of the means for all samples of size 40 can be calculated as follows:

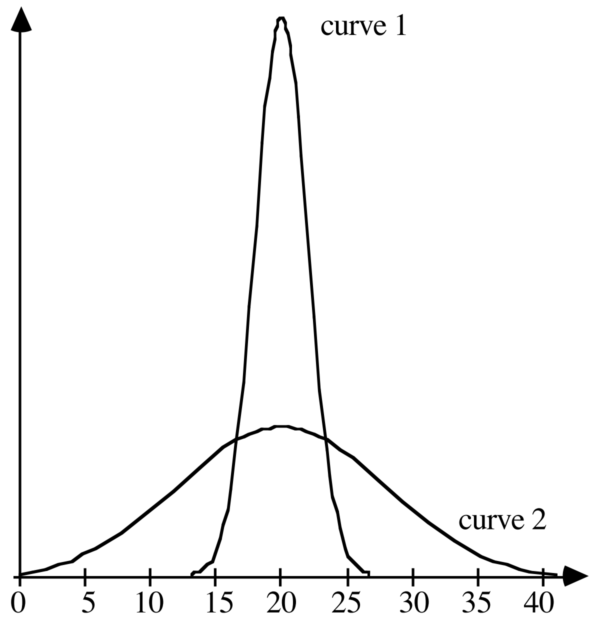
$$\sigma_{\bar{x}} = \frac{5}{\sqrt{40}} \approx 0.79$$

f. How does the sample size n affect the shape of the distribution of all possible sample means? Explain why this occurs.

g. When sampling a population, how can a researcher be reasonably sure that the sample mean is a good estimate of the population mean?

Assignment

2.1 The figure below shows two normal curves. One curve represents the distribution of a characteristic in a population. The other represents the distribution of sample means for that characteristic in the same population, using a sample size of 16.



a. Which curve represents the distribution of sample means? Defend your choice, including an explanation for the difference in the shapes of the two curves.

b. Use the appropriate curve to estimate the mean and standard deviation of the population.

c. Use your responses from Part **b** to estimate the mean and standard deviation of the sample means.

2.2 A family of six has the following heights (in centimeters): 145, 156, 163, 170, 174, and 188.

a. Find μ for this population.

b. Use combinations to determine the number of samples of two heights that can be taken from this population.

c. List all the possible samples of two heights from the population.

d. Find the mean of each sample in Part **c**.

e. Find the mean of five randomly chosen sample means from Part **d**.

f. Find the mean of all the sample means from Part **d**.

g. Compare the values obtained in Parts **a**, **e**, and **f** and write a summary of your findings.

2.3 **a.** Randomly generate 10 whole numbers from 1 to 9. Calculate the mean of these numbers.

b. Repeat Part **a** 19 more times, creating a population of 200 numbers from 1 to 9, and a group of 20 sample means.

c. Create a relative frequency histogram of the population of 200 numbers and describe the shape of the graph.

d. Create a relative frequency histogram of the 20 sample means and describe the shape of the graph.

e. How does the distribution of sample means compare to the distribution of the population from which the samples were taken? Use the central limit theorem to explain why the difference occurs.

2.4 A bottling plant fills bottles with soda. The volumes of the population of filled bottles are normally distributed, with a mean of 355 mL and a standard deviation of 2 mL. As part of the quality control process, a sample of four bottles is selected every hour. A technician records the mean volume of each sample. What are the mean and standard deviation of these sample means?

* * * * *

2.5 Consider a population consisting of the whole numbers from 1 to 99.

a. Determine the standard deviation of this population (σ).

b. Determine the standard deviation of the sample means ($\sigma_{\bar{x}}$) for samples of size 30 taken from this population.

c. How does the standard deviation of the population differ from the standard deviation of the sample means? Explain why you would expect this difference to occur.

2.6 The Sure Grip Tire Company manufactures motorcycle tires. The life spans of a population of its tires are normally distributed with a mean of 85,000 km and a standard deviation of 3,750 km.

a. What is the standard deviation of the sample means for samples of size 100 taken from this population?

b. How could the company decrease the size of the standard deviation of the sample means?

* * * * * * * * * *

Activity 3

According to the central limit theorem, even if a population is not normally distributed, the distribution of sample means can often be approximated reasonably well by a normal curve. This is one of the most useful facts in statistics.

Quality-control engineers, for example, frequently measure quality in terms of a product's mean life. Whenever possible, they model the results of their experiments with normal curves. When working with normal curves, they can express the proportion of the data located within a specific interval as a percentage. In Figure **5**, for example, the area under the curve that corresponds to the proportion of data in the interval $[a,b]$ is 55%. In this activity, you explore some of the properties that make normal curves so useful.

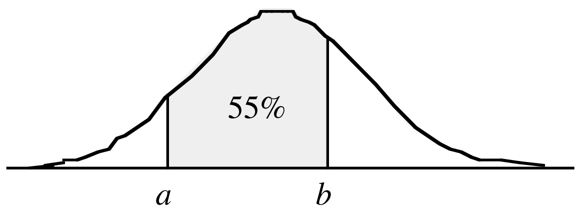


Figure 5: Percentage of area corresponding to [*a*, *b*]

Exploration

In this exploration, you use the set of whole numbers from 1 to 999 as a model population ($\mu = 500$ and $\sigma = 288$).

a. Calculate the standard deviation of the sample means ($\sigma_{\bar{x}}$) for samples of size 30 taken from this population.

b. Sketch a normal curve that models the distribution of sample means for samples of size 30.

c. Label the mean (μ) on the *x*-axis, as well as the values that are 1, 2, and 3 standard deviations from μ.

d.
1. Use technology to generate 50 random samples of size 30 from the population of whole numbers from 1 to 999.
2. Calculate the means of the 50 samples.

e. Determine the number of sample means that falls within each of the following intervals under the normal curve from Part **c**.
1. $[\mu - 1\sigma_{\bar{x}},\ \mu + 1\sigma_{\bar{x}}]$
2. $[\mu - 2\sigma_{\bar{x}},\ \mu + 2\sigma_{\bar{x}}]$
3. $[\mu - 3\sigma_{\bar{x}},\ \mu + 3\sigma_{\bar{x}}]$

f. Calculate the percentage of sample means that falls within each of the intervals described in Part **e**.

g. Collect the sample means obtained by the entire class. Determine the percentage of the class data that falls within each of the intervals described in Part **e**.

Mathematics Note

The **68–95–99.7 rule** states that approximately 68% of the total area between the normal curve and the *x*-axis lies within 1 standard deviation of the mean, 95% lies within 2 standard deviations of the mean, and 99.7% lies within 3 standard deviations of the mean. This rule is illustrated in Figure **6**.

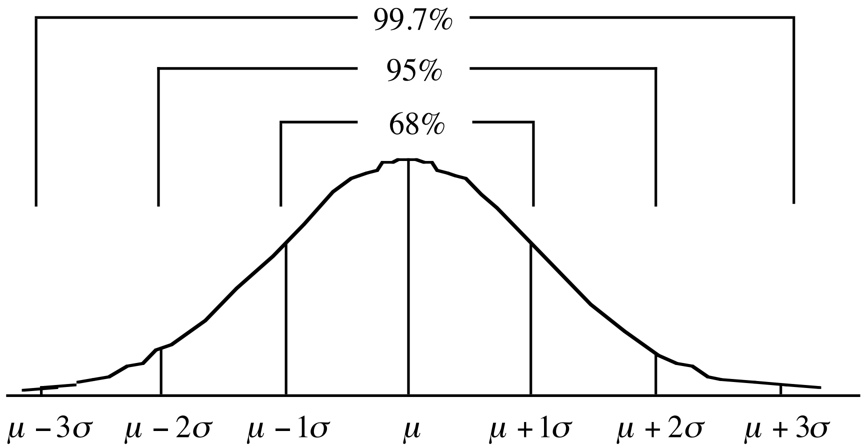


Figure 6: A normal curve and the 68–95–99.7 rule

For example, if the population mean is 100 and the standard deviation is 10, then you might expect about 68% of the sample means to lie between 90 and 110, 95% of the sample means to lie between 80 and 120, and 99.7% of the sample means to lie between 70 and 130.

Discussion

a. How did the percentages you obtained for the distribution of sample means compare to those described in the mathematics note?

b. How did the class percentages for the distribution of sample means compare to those in the mathematics note?

c. Explain why you might expect some differences between the percentages obtained in the exploration and the percentages given in the mathematics note.

d. How would you modify the exploration to obtain more accurate percentages?

Assignment

3.1 While researching automobile tires, you find that the mean life for samples of one particular brand is 60,000 km with a standard deviation of 1200 km. Assume that the life spans of tires in this population are normally distributed.

a. What percentage of the population would you expect to have life spans in the interval [58,800, 61,200]? Explain your response.

b. What are the upper and lower bounds of the interval within which you would expect the life spans of 95% of these tires to be found? Describe how you determined this interval.

3.2 Samples of a certain brand of light bulbs have a mean life of 1015 hr with a standard deviation of 75 hr.

a. What percentage of these bulbs would you expect to last longer than 1090 hr?

b. If a company purchases 1000 of these bulbs for use in its manufacturing plant, how many of them would you expect to burn out after less than 865 hr of use?

3.3 A quality control engineer who tests automobile seat belts selects a random sample of 36 men, aged 18 to 74, and measures their masses in kilograms. These masses are shown in the following table.

67.7	81.0	50.6	74.4	94.8	65.8
79.7	93.8	56.2	72.8	89.6	89.3
63.6	76.3	85.2	88.4	69.6	77.6
64.9	69.1	55.2	80.7	63.4	59.4
53.6	59.2	96.4	57.4	103.4	90.2
103.8	81.4	65.6	101.2	87.5	65.6

a. The masses of the population from which the engineer selected the sample are normally distributed with a mean of 75 kg and a standard deviation of 14.9 kg. Determine an interval that the engineer could expect to contain 68% of the data.

b. Find the actual percentage of data that lies within the interval you determined in Part **a**.

c. Explain why the percentage calculated in Part **b** might differ from 68%.

3.4 The quality control engineer described in Problem **3.3** decides to collect data from 19 more random samples of men (20 in all). The means of the samples (in kilograms) are shown in the table below.

76.0	83.1	75.8	82.1	68.4
76.1	80.6	77.4	79.6	75.4
90.0	68.9	85.4	80.5	82.0
81.2	73.4	62.3	83.2	82.5

a. Find the percentage of sample means that lie within the interval calculated in Problem **3.3a**.

b. Compare the percentage of sample means that lie within the interval to the percentage of masses from the single sample in Problem **3.3**. Explain any differences you observe.

* * * * *

3.5 A test was given to 3,000,000 people. The scores on the test were normally distributed, with a mean of 900 and a standard deviation of 212.

a. Determine the interval that contains 99.7% of the scores.

b. What percentage of the population received scores less than or equal to 1324?

c. What percentage of the population received scores less than or equal to 688?

3.6 The supervisor of the waiters and waitresses at El Burrito restaurant noticed that the amount of tips reported per shift seemed to be normally distributed with a mean of $27.35 and a standard deviation of $10.17.

a. Find the interval that contains 95% of the reported tips.

b. Find the interval that contains 99.7% of the reported tips.

c. What percentage of the staff make more than $17.18 in tips per shift?

d. From your experience with restaurants, would you have expected the amount of tips reported per shift to be normally distributed? Explain your response.

* * * * * * * * * *

Activity 4

When conducting statistical studies, researchers usually are concerned with large populations in which the mean value of a characteristic is unknown. In these situations, an estimate of the population mean is made using the statistics from a single sample. In this activity, you examine how confident researchers can be in the accuracy of their estimates.

Exploration

In this exploration, you again use the numbers from 1 to 999 as a model population ($\mu = 500$ and $\sigma = 288$).

a. Select a random sample of 30 numbers from the population and calculate $\bar{x}$.

b. Determine s, the standard deviation of this sample.

c. In most real-world situations, the mean of the population is unknown. In such cases, the sample mean ($\bar{x}$) is used to approximate the population mean (μ). In a similar manner, the sample standard deviation (s) is used to estimate the population standard deviation (σ).

1. Use the value for s from Part **b** to determine a value for the standard deviation of the sample means. In other words, estimate $\sigma_{\bar{x}}$ by substituting s for σ in the formula below, where n is the sample size:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

2. You can now use the mean of the sample ($\bar{x}$)and standard deviation of the sample means $(\sigma\ \bar{x})$to describe an interval that may contain the population mean. Create an interval that describes values that are no more than 2 standard deviations of the sample means on either side of $\bar{x}$.

d. Determine if the interval you created in Part **c** contains μ.

e. Repeat Parts **a–d** 49 more times (for a total of 50 samples). Record the number of times that the interval generated did not contain the population mean.

Discussion

a. What percentage of the time did the interval created in the exploration not contain the population mean?

b. The intervals you created in Part **c** of the exploration are known as **95% confidence intervals**. What would you expect to be true of these intervals?

c. Do your results in the exploration support the 95% figure?

d. How would you modify the process described in Part **c** of the exploration to obtain a 68% confidence interval?

e. How would you modify the sampling procedure to obtain a narrower 95% confidence interval? Explain your response.

Mathematics Note

A **confidence interval** for a parameter is an interval of numbers in which one would expect to find the value of that parameter. Every confidence interval has two aspects: an interval determined by the statistics collected from a random sample and a confidence level that gives the probability that the interval includes the parameter.

For example, a 95% confidence interval is generated by a process that results in an interval in which the probability that the parameter lies in that interval is 95%. In other words, you would expect 95% of the intervals produced by this process to contain the parameter, while 5% of the intervals would not.

The mean of all the sample means of a given sample size equals the mean of the population μ. Since sample means are normally distributed, the 68–95–99.7 rule can be applied. For example, the mean $\bar{x}$ of any one sample will fall within 2 standard deviations of μ 95% of the time. This fact also indicates a 95% probability that μ is within 2 standard deviations of $\bar{x}$. This 95% confidence interval can be represented algebraically as shown below:

$$\bar{x} - 2\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + 2\left(\frac{\sigma}{\sqrt{n}}\right)$$

Since the population standard deviation usually is unknown, it is necessary to use the sample standard deviation s as an estimate of σ. In order to ensure that s properly approximates σ, the sample size n must be at least 30.

For example, imagine that a biologist selects a random sample of 100 fish from a lake. The mean length of the fish in the sample is 11 cm, with a standard deviation of 2.5 cm. To determine a 95% confidence interval, the biologist substitutes 11 for $\bar{x}$, 2.5 for s, and 100 for n as shown below:

$$11 - 2\left(\frac{2.5}{\sqrt{100}}\right) \le \mu \le 11 + 2\left(\frac{2.5}{\sqrt{100}}\right)$$

$$10.5 \le \mu \le 11.5$$

The biologist can then declare with 95% confidence that the mean length of fish in the population is in the interval [10.5, 11.5].

Assignment

4.1 **a.** Imagine that you have selected two random samples of the same size from the same population. Using the statistics from the two samples, you then determine two 95% confidence intervals for the population mean. Would you expect the two confidence intervals to be the same? Explain your response.

b. If you took 20 samples of the same size from the same population and determined 20 corresponding 95% confidence intervals, how many of them would you expect to contain the population mean? Explain your response.

4.2 **a.** For a given sample, would a 99.7% confidence interval be larger or smaller than a 95% confidence interval? Explain your response.

b. Write an algebraic representation of a 99.7% confidence interval given the sample mean ($\overline{x}$), sample standard deviation (s), and sample size (n).

4.3 A fish food manufacturer is developing a product to increase first-year growth in trout. After a year on this experimental diet, a random sample of 100 trout revealed a mean gain in mass of 84 g with a standard deviation of 14 g.

a. Use s to approximate the standard deviation of all possible sample means for $n = 100$.

b. Determine a 95% confidence interval for the mean gain in mass of trout fed the experimental diet.

c. Describe the meaning of the confidence interval in Part **b**.

d. How might the standard deviation of sample means be affected if the sample size were increased to 400 trout?

e. Why might a smaller standard deviation of sample means help the company market its fish diet?

4.4 A car manufacturer claims that its new model consumes fuel at a rate of 14 km/L. To verify this claim, the quality control engineers at a competing company selected a sample of 16 cars. Their tests yielded a mean of 13.5 km/L and a standard deviation of 1.6 km/L.

a. Create a 68% confidence interval for the rate of fuel consumption and determine if it contains the figure claimed by the manufacturer.

b. Create a 95% confidence interval for the rate of fuel consumption and determine if it contains the figure claimed by the manufacturer.

c. Do you believe the manufacturer's claim? Write a paragraph explaining how you reached your conclusion.

4.5 An advertising agency is investigating the number of hours that the residents of a particular city spend watching television each day. As part of their study, they surveyed a random sample of city households. The results of the survey are shown in the frequency table below.

Number of Hours Watching Television Per Day	**Frequency**
1	4
2	5
3	10
4	6
5	4
6	2
7	1
8	1

a. Construct a 95% confidence interval for the mean number of hours that city households spend watching television per day.

b. What does your response to Part **a** indicate about television viewing?

4.6 A 95% confidence interval for the mean life (in hours) of a particular brand of batteries is $410 \le \mu \le 450$.

a. Determine a 99.7% confidence interval for the mean life of these batteries.

b. What does your response to Part **a** indicate about battery life?

4.7 What factors should investigators consider when deciding whether to use a 68%, a 95%, or a 99.7% confidence interval?

* * * * *

4.8 A soft-drink company is monitoring the performance of its bottling equipment. According to the label, each bottle should contain 340 mL of soda. A quality-control specialist selects a random sample of 100 bottles each day for 5 days and measures their volumes. The table below displays the values for $\bar{x}$ and s for each day.

Day	$\bar{x}$ **(mL)**	s **(mL)**
1	340.8	2.6
2	340.2	2.1
3	340.9	1.7
4	340.4	2.3
5	340.1	0.9

a. Using the data for day 1, construct a 95% confidence interval for the mean volume of soda per bottle.

b. Using the data for day 2, construct a 95% confidence interval for the mean volume of soda per bottle.

c. **1.** Compare the two confidence intervals from Parts **a** and **b**.

2. How many different 95% confidence intervals would you get if you calculated one for each day of available data?

d. Is there any guarantee that any of the 95% confidence intervals obtained from the data will contain the actual mean volume per bottle? Explain your response.

e. Do you think that the company's labels accurately describe the milliliters of soda in a bottle? Write a paragraph to explain your conclusion.

* * * * * * * * * *

Research Project

Select a population and choose one of its characteristics to study. For example, if you select the students at your school as a population, you might wish to examine the mean number of hours spent studying per week.

Develop a sampling method that will allow you to determine a reasonably good estimate of the mean of this characteristic. After conducting your study, write a report that includes the following:

- descriptions of the population you surveyed, the characteristic you examined, and your sampling method
- the data you collected as well as the statistics your data generated, including a confidence interval
- a statement summarizing the conclusions you drew from your analysis
- any suggestions or recommendations that you believe your study supports.

Summary Assessment

Imagine that you are a journalist for the local newspaper. Your editor has requested that you write an article that uses statistics to describe one characteristic of a population. Your article must include two parts.

The first part of your article should describe the population, the characteristic you measured, and the sampling method used. It also should include the sample data you obtained.

To write this portion of your article, complete the following steps.

- Create a hypothetical population with some measurable characteristic.
- Design a method to sample this population.
- Create a simulation that models both your population and your sampling method.
- Use your simulation to generate some sample data.

The second part of your article should report on the results of your study. Include an estimate of the mean value of the characteristic for the population, and support your estimate with a discussion of confidence intervals and the normal distribution of sample means. To illustrate the mathematics involved in your analysis, you also should include the values for $\bar{x}$ and $\sigma_{\bar{x}}$ generated from your data, as well as a 95% confidence interval for μ.

Your article should conclude with any recommendations or suggestions that you feel your study supports.

Module Summary

- The mean value for a population, or **population mean**, is denoted by the Greek letter μ (mu).
- The **population standard deviation** is denoted by the Greek letter σ (sigma). It can be calculated using the following formula:

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}}$$

  where the population has *N* members represented by $x_1, x_2, \ldots, x_N$.
- The mean value for a sample, or **sample mean**, is denoted by $\bar{x}$ (read "x-bar").
- The **sample standard deviation**, denoted by *s*, can be calculated as follows:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}}$$

  where the sample has *n* members represented by $x_1, x_2, \ldots, x_n$.
- A **normal distribution** is a continuous probability distribution. The graph of a normal distribution is symmetric about the mean and tapers to the left and right like a bell. The curve that describes the shape of the graph is the **normal curve**. As in all continuous probability distributions, the total area between the *x*-axis and a normal curve is 1.
- Although all normal curves have the same general shape, the width of any particular curve depends on the standard deviation (σ) of the distribution that the curve models. Since the population mean (μ) is located at the point where the curve's axis of symmetry intersects the *x*-axis, the position of the curve along the *x*-axis depends on the value of μ.
- The **law of large numbers** states that, for very large sample sizes, there is a greater probability that the sample mean is close to the population mean.
- The **central limit theorem** states that, even if the population from which samples are taken is not normally distributed, the distribution of the means of all possible samples of the same size will be approximately normal.

  This approximation becomes more accurate as the sample size *n* increases. For $n \geq 30$, the distribution of sample means can be modeled reasonably well by a normal curve. This requirement is not necessary if the population from which samples are taken is normally distributed.
- The mean of the distribution of all sample means for a given sample size equals μ, the population mean.

- The **standard deviation of all sample means**, denoted by $\sigma_{\bar{x}}$, can be calculated using the following formula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

  where σ is the standard deviation of the population and n is the sample size.

- The **68–95–99.7 rule** states that approximately 68% of the total area between the normal curve and the *x*-axis lies within 1 standard deviation of the mean, 95% lies within 2 standard deviations of the mean, and 99.7% lies within 3 standard deviations of the mean.

- A **confidence interval** for a parameter is an interval of numbers in which one would expect to find the value of that parameter. Every confidence interval has two aspects: an interval determined by the statistics collected from a random sample and a confidence level that gives the probability that the interval includes the parameter.

Selected References

Brase, C. P., and C. H. Brase. *Understandable Statistics: Concepts and Methods*. Lexington, MA: D. C. Heath, 1991.

Miller, I., J. E. Freund, and R. A. Johnson. *Probability and Statistics for Engineers*. Princeton, NJ: Prentice-Hall, 1990.

Moore, D. S. *Statistics: Concepts and Controversies*. San Francisco, CA: W. H. Freeman and Co., 1979.

Neter, J., W. Wasserman, and G. A. Whitmore. *Applied Statistics*. Boston, MA: Allyn and Bacon, 1993.

Rowntree, D. *Statistics Without Tears: A Primer for Non-Mathematicians*. New York: Scribner's, 1981.