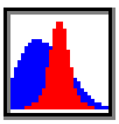
CHAPTER 7



Visualizing the Central Limit Theorem

CONCEPTS

- Central Limit Theorem, Population Being Sampled, Sampling Distribution, Population Mean, Standard Error, Standard Deviation, Asymptotic

OBJECTIVES

- Recognize the difference between the population being sampled and the sampling distribution of the mean

- Comprehend the importance of the Central Limit Theorem

- Discern the difference between the standard error of the sampling distribution and the standard deviation of the population

- Understand the role of sample size in determining the standard error of the sampling distribution

Overview of Concepts

Pierre Simon Laplace's (1749–1827) major contribution to probability theory is now called the **Central Limit Theorem**, where the term *central* means fundamental rather than middle or center. His work was a generalization of the work of Abraham De Moivre. Whereas De Moivre (1667–1754) had shown that if the number of trials is very large, a symmetric binomial distribution becomes what we now call the normal distribution, Laplace showed that the sum or mean of independently drawn random variables will, if the sample size is large, be approximately normally distributed. Over the years this was further generalized and regularity conditions were added until the final form we use today was produced by Lindberg in 1922.

The Central Limit Theorem can be formally stated as follows: If $X_1, X_2, \ldots, X_n$ are identically distributed independent random variables from any population with finite mean μ and finite variance σ^2, then the sample mean $\bar{X} = (x_1 + x_2 + \ldots + x_n) / n$ will have a sampling distribution with mean μ and variance σ^2 / n. Further, the sampling distribution of $\bar{X}$ is asymptotically normally distributed.

This theorem has several important aspects. First, the **population being sampled** must have both a **population mean** (μ) and a **standard deviation** (σ). While the vast majority of distributions satisfy this requirement, there are distributions that lack a standard deviation, and some that have neither a standard deviation nor a mean. For example, a Student's t distribution with 2 degrees of freedom doesn't have a standard deviation (if you have not studied this distribution, you will later in the course), while a Student's t with 1 degree of freedom doesn't have a mean or a standard deviation. Second, all of the random variables $X_1, X_2, \ldots, X_n$ must be *independently* drawn from the same population. Therefore, drawing one random variable will not affect the chances of drawing any other random variable. This would be violated, for example, if you sampled from a hypergeometric distribution. Third, the sample mean has its own distribution called a **sampling distribution**. The mean of the sampling distribution is μ, the same as the mean of the population being sampled. The **standard error** (denoted $\sigma_{\bar{x}}$) of the sampling distribution is $\sigma / \sqrt{n}$ and is the square root of the variance of the sampling distribution. Note that the standard error will get smaller as your sample size n increases. Fourth, if the sample size is large enough, then the sampling distribution is approximately normally distributed, regardless of the population's shape. Since this will always be true only as $n \to \infty$, this is called an **asymptotic** result.

An intuitive explanation of the Central Limit Theorem is possible. Assume that you draw a sample of n observations from a population and calculate the sample mean. You then draw a new sample and calculate a new sample mean. If you redo this process many times, the average of these sample means will be near μ, and their standard error will be near $\sigma / \sqrt{n}$. Further, if n is large enough and you made a dot plot or histogram of all of the sample means, they would approximately form a normal distribution.

This is a very important theorem. It tells a statistician that even if the underlying population is skewed, or bimodal, the sample means will be approximately normally distributed if a large enough sample size is used. It also says that regardless of the size of the sample, the sample mean has the same mean as the original population and its standard error is much smaller than the standard deviation of the population sampled. This result will be invaluable when you learn to make inferences about a population.

Illustration of Concepts

A coal-fired electric power plant emits sulfur dioxide (SO_2). The power plant continuously measures these emissions and records its average emissions daily. To be in compliance with the EPA, a plant's daily emissions can exceed 140 parts per billion (ppb) only one day per year and must average below 30 ppb for the entire year. Twenty samples are taken. Is the power plant currently in compliance? Will the power plant be in compliance at the end of the year?

Below is a histogram of a sample of SO_2 emissions. The population being sampled has been superimposed as a line (this is drawn for your benefit but is not known to the company). It is clear from the histogram that the plant does not have any emissions over 140 ppb. However, it is not clear whether the plant will meet the yearly requirement of averaging under 30 ppb. Based on the 20 days sampled, the sample mean is 32 ppb. Although the **mean** of the **population being sampled** is unknown, the company's statistician knows, based on experience with the technology, that the population is positively skewed and has a **standard deviation** of 16. Therefore she poses the following question: "Is this sample mean consistent with the belief that the true mean is less than 30 ppb?"

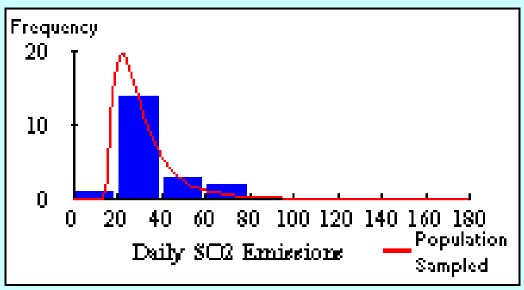


Figure 1: Histogram of Sulphur Dioxide Emissions

In order to answer this question she used the **Central Limit Theorem** to find out what the mean should be to make it very unlikely (since the EPA fine for noncompliance is very large) that the sample mean would be greater than 30 at the end of the year. The Central Limit Theorem provides the **asymptotic** result that if $n \to \infty$, then the **sampling distribution** is normally distributed. At the end of the year the company will have 365 samples. This is large enough to know that the sampling distribution is normally distributed. In addition, since the standard deviation is 16, the **standard error** of the sample mean at the end of the year is 0.8375 (16 divided by the square root of 365). Therefore, if the population mean is no more than 27.9 ($= 30 - 2.5 \times 0.8375$), there is less than a 0.62% chance of being fined.

The question then becomes, "Is a sample mean of 32, after 20 observations, consistent with a population mean of 27.9?" To make this calculation, the statistician assumes that the sample of 20 is large enough so that the sampling distribution is approximately normally distributed (she knows, given the modest degree of skewness in the population being sampled, that this is not a bad assumption). The standard error of the sample mean with 20 observations is 3.578 (16 divided by the square root of 20). Therefore the probability that $\bar{X} \geq 32$ is

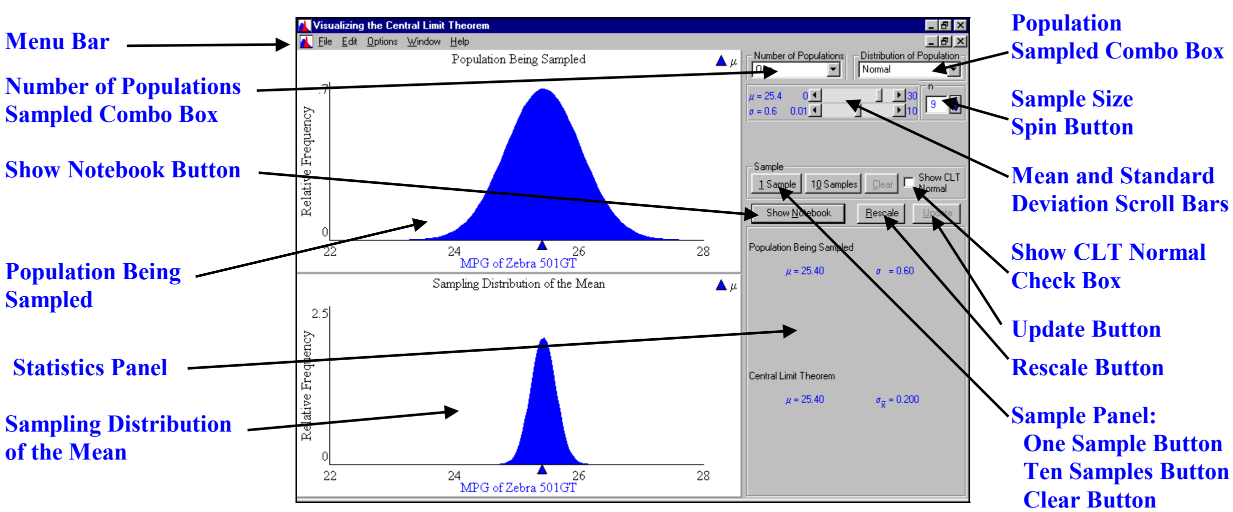
$\Pr(\bar{X} \geq 32 \mid \mu = 27.9 \text{ and } \sigma_{\bar{x}} = 3.578) = \Pr(Z \geq (32-27.9)/3.578) = \Pr(Z \geq 1.14) = 0.1271$.

Hence, there is a 13% chance that the power plant's actual mean is 27.9 ppb or less. Given the size of the fine if the power plant does not meet the 30 ppb requirement, the company should use the 11 months still remaining to readjust their scrubbers and other pollution abatement equipment to increase their chances of meeting the EPA's yearly requirement.

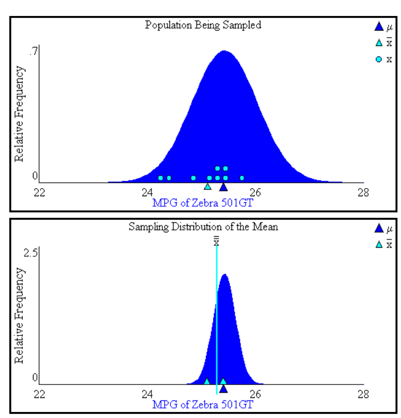
Orientation to Basic Features

This module allows you to draw a sample from a normal, uniform, or skewed distribution, in order to demonstrate the similarity between the Central Limit Theorem's predictions and the sample results. You can also sample from two populations simultaneously to compare different population parameters or different sample sizes.

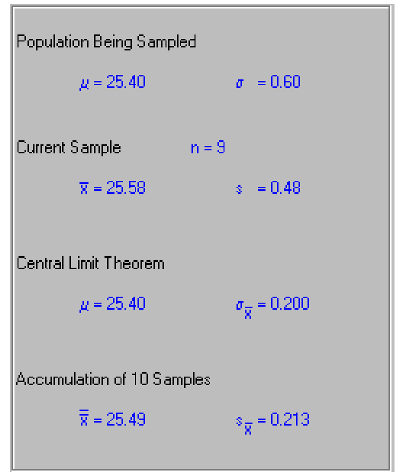
1. **Opening Screen**
   Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. Read the questions and then click the Concepts tab to see the concepts that you will learn. Click on the Scenarios tab. Select Normal Distribution from the table of choices. Select a scenario, read it, and press OK. The upper left of the screen shows the population being sampled. The Control Panel is to its right. On the lower left of the screen is the sampling distribution of the mean as stated in the Central Limit Theorem (assuming a large sample size). To the right of the sampling distribution is the statistics panel containing the population and sampling distribution statistics. The population mean is shown as a fulcrum below each distribution.



2. **Control Panel**
   a. Click the One Sample button in the Sample panel. A dot plot of the sample is superimposed on the population being sampled. A small fulcrum showing the sample mean appears below the dot plot (the large fulcrum is the population mean). That sample mean is placed on the sampling distribution. Push the One Sample button again to generate a second sample and dot plot. The new sample mean is shown on the top display and is placed on the sampling distribution. The mean of these two sample means is shown as a vertical line on the sampling distribution. The fulcrum is the population mean.



Sample statistics appear in the statistics panel to the right. Push the Clear button in the Sample panel to clear the dot plot and sample statistics from the displays. Push the Ten Samples button to draw 10 samples and display them on the dot plot one after the other. Press the Stop Sampling button to quit after the last sample is displayed. Press the Skip to End button if you want to avoid redrawing the displays after each sample is drawn. The final result of taking the 10 samples is still displayed.

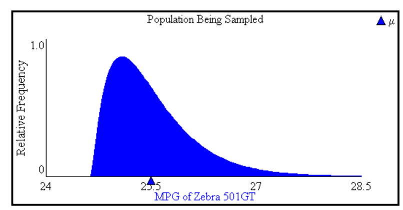


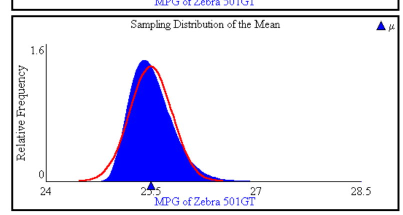
b. Click the n spin button to change the sample size. Press the flashing Update button to redraw the display. Note that the sampling distribution changes.

c. Click the μ or σ scroll bars to change the mean or standard deviation. Press the flashing Update button to redraw the display with a new mean or standard deviation. Note how both distributions change. (Since you have changed the population parameters, the scenario may no longer have meaning.)

d. Push the Rescale button to rescale both distributions. A warning message appears if the distributions disappear from the screen.

e. Click on the population sampled combo box and select Skewed Population (since you have changed the population, the scenario may no longer have meaning). The population being sampled changes, as does the sampling distribution (illustrated on the right). The Central Limit Theorem says that sampling distribution of the mean will be normal *if* the sample size is large enough (its mean and standard error are shown on the Statistics panel). Click Show CLT Normal to superimpose the normal distribution predicted by the Central Limit Theorem on the sampling distribution in the lower diagram. In this illustration, the sampling distribution is not exactly normal.





3. **Copying a Display**

Click on the display you wish to copy. Its window title will be highlighted. Select Copy from the Edit menu (on the menu bar at the top of the screen) or Ctrl-C to copy the display. It can then be pasted into other applications, such as Word or WordPerfect, so it can be printed.

4. **Help**

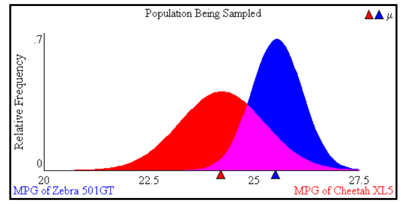
Click on Help on the menu bar at the top of the screen. Search for Help lets you search a topic index, Contents shows a table of contents for this module, Using Help gives instructions on how to use Help, and About gives licensing and copyright information about this *Visual Statistics* module.

5. **Exit**

Close the module by selecting Exit in the File menu (or click ☒ in the upper right-hand corner of the window). You will be returned to the *Visual Statistics* main menu.

Orientation to Additional Features

1. **Two Populations**
Click on the number of populations sampled combo box and select Two Populations. Two populations can now be sampled simultaneously. This can be useful to compare distributions with different means, standard deviations, or sample sizes. The area of overlap of the red and blue distributions is shown in magenta.



2. **Do-It-Yourself**
Push the Show Notebook button and select the Do-It-Yourself tab. Press OK to return to the main screen with the illustrative scenario labels removed from the graphs. Use the μ and σ scroll bars to create your own examples.

3. **Options**
Five sets of options are available from the Options menu on the menu bar.
   a. Select Change Title if you wish to retitle the current display.
   b. Select Auto Update if you wish to have the displays automatically updated when a Control Panel change is made.
   c. Select Auto Rescale if you wish to have the displays automatically rescaled when the displays are updated. With this option, graphs often look unchanged, but the scale on the axis has changed.
   d. Select Full Window Graph from the Options menu. This extends your graph to the full width of your computer screen. Deselect Hide Controls in the Options menu to make the controls reappear.

4. **Second Display**
Select either Copy Default Window or Copy Current Window from the Windows menu on the menu bar. This creates a second graph that you can Tile or Cascade (from the Windows menu). Use the Change Title option (see 3a above) to retitle the display. This feature is useful to label different examples that are displayed simultaneously.

Basic Learning Exercises Name ________________

Sampling a Normal Population

Press the Show Notebook button, select the Scenarios tab, click on Normal Distribution, and select Fuel Economy. Read the scenario and click on OK.

1. What is the mean and standard deviation of the population being sampled? What is the mean and standard error of the sampling distribution? What is the relationship between the standard deviation of the population being sampled and the sampling distribution? If you aren't sure use the Help option on the menu bar.

   Population Being Sampled: Mean ___________ Standard Deviation ___________
   Sampling Distribution: Mean ___________ Standard Error ___________

2. Push the One Sample button on the Sample panel. Do all of the *individual* observations on mileage meet the company's expectation of at least 25 mpg? On *average*, do the cars meet the company's expectations? To which diagram (upper or lower) did you refer to answer these questions? If there is a difference in your answers to these questions, explain.

3. If the company were to advertise average mileage of 25 mpg, would this be misleading to its customers? Explain.

4. Increase the sample size to 25 by using the n spin button or by typing directly into the box and pressing the Enter key. Click Update. What is the standard deviation of the population? What is the standard error of the sampling distribution? Why did the latter change? How did the lower diagram's appearance change?

   Standard Deviation ___________ Standard Error ___________

5. If you wanted a standard error of 0.06 for the sampling distribution, what sample size would you need? Explain.

6. Give a formal statement of the Central Limit Theorem. In your own words, what does it mean? **Hint:** Use the Help feature if you do not remember the CLT.

7. How does exercise 5 use the Central Limit Theorem?

8. Reduce n to 2 using its spin button. Press the Ten Samples button five times. Is the sampling distribution formed by the dot plot normally distributed? Is so, why; if not, why not?

9. Press the Show Notebook button and select the Two Different Fruit Sorters scenario. Read the scenario and click OK. Describe the populations that you see.

10. Answer the two questions in the scenario (click the Show Notebook button if you need to re-read the scenario). How does the Central Limit Theorem help you answer the questions posed in the scenario?

Intermediate Learning Exercises Name ________________

Sampling a Nonnormal Population

Press the Show Notebook button, select the Scenario tab, and select Other Distributions. Select the Pizza Delivery scenario. Read the scenario and press the OK button.

11. Why would the time it takes to deliver a pizza have a skewed distribution like the one shown?

12. Select Show CLT Normal on the Control Panel. Press the Ten Samples button four times. This generates 40 sample means and displays them on the sampling distribution. Does the experimental sampling distribution you created look like the normal sampling distribution based on the CLT that is shown? Why or why not? Is this result consistent with the Central Limit Theorem?

13. If you increased your sample size to 20, would you expect the experimental sampling distribution to be similar to the normal sampling distribution? *Answer this exercise without using the computer.*

14. Increase your sample size to 20 by using the n spin button or entering the number directly into the spin button box and pressing Enter on your computer. Press the Ten Samples button four times. Did you answer exercise 13 correctly? If not, re-answer that same exercise below.

15. If you had a very skewed distribution, would a sample size of 20 always be large enough to generate a normal sampling distribution? **Hint**: Change the distribution to Very Skewed.

16. Push the Show Notebook button and select the Two Truck Drivers scenario. Read the scenario and push the OK button. Answer the scenario's questions. How does this scenario illustrate the Central Limit Theorem?

17. Push the Show Notebook button and select the Random Number scenario. Read the scenario and push the OK button. Answer the scenario's questions. How does this scenario illustrate the Central Limit Theorem?

18. Reduce n to 2 using its spin button. Press the Ten Samples button five times. Is the sampling distribution formed by the dot plot normally distributed? If so, why? If not, why not?

19. Increase n to 6. Press the Ten Samples button five times. Is the sampling distribution formed by the dot plot normally distributed? If so, why? If not, why not?

20. Is it easy or difficult to see the exact shape of the sampling distribution by looking at the dot plot of the sample means (the little triangles)? Explain.

Advanced Learning Exercises Name ________________

An Unbiased Estimator for μ

Push the Show Notebook button, select the Do-It-Yourself tab and press OK. Use the list boxes at the top of the Control Panel to show One Population and Skewed Population. Set σ to 20. Set μ to any value you wish. Push the flashing Update button.

21. Set sample size to 2 using the n spin button or entering the number in the spin button box. Press the One Sample button twice. The sample means from each draw are shown as a dot plot on top of the sampling distribution. Notice that a vertical line is drawn through the sampling distribution. This is the average of the two sample means. It is denoted as $\overline{\overline{X}}$. The average should be located halfway between the two sample means (because of the nature of the dot plot of the two sample means this sometimes appears to be incorrect, but it is not). Where is $\overline{\overline{X}}$ in relation to the population mean μ (the fulcrum below the sampling distribution)?

22. Press the One Sample button again. How did the position of $\overline{\overline{X}}$ change?

23. Press the One Sample button three more times watching what happens to the position of $\overline{\overline{X}}$ after each new sample is drawn. Do you notice a general pattern? If you do, explain what you have noticed. If you don't, continue to press the One Sample button until a pattern emerges, and then describe the pattern you observed. (Sometimes because of sampling variation a pattern was not evident. If this occurs after 10 samples are drawn, press the Clear button on the Sample panel and repeat exercises 21–22.)

24. Verify this pattern by pressing the Ten Samples button. Press the Ten Samples button a second time. Press the Ten Samples button a third time. What pattern did you observe?

25. If the average of many estimates equals the parameter's value, then the estimator provides an unbiased estimate for the parameter. *Therefore, the sample mean is an unbiased estimator for μ.* If this property is *only* true for large sample sizes, then the property is called asymptotically unbiased. Increase the sample size to 25. What do you observe; does the same pattern hold? Are there any modifications to your previous observation that you could make as a result of increasing the sample size?

A Consistent Estimator for μ

26. An estimator is called a consistent estimator for a parameter if the estimator is either an unbiased or an asymptotically unbiased estimator for the parameter *and* if as the sample size increases, the standard error of the sampling distribution gets smaller. What is the standard error if sample size is 4? 16? 64? 256? Is the sample mean a consistent estimator for μ?
    Note: Advanced students may recognize that this is not the correct formal definition of consistency. This type of consistency is sometimes called Mean Squared Error Consistency. However, for practical purposes, it is an appropriate substitute.

    Standard error if n = 4 __________ Standard error if n = 16 _________
    Standard error if n = 64 __________ Standard error if n = 256 _________

Individual Learning Projects

Write a report on one of the three topics listed below. Use the cut-and-paste facilities of the module to place the appropriate graphs in your report.

1. Investigate the normality of the sampling distribution. Set your sample size to 2 and draw 50 samples. Are your sample means normally distributed? If not, double your sample size and draw 50 more samples. Continue this process until your plot of sample means is normally distributed. What sample size was needed? Repeat the process for the other two populations. Write a report on what you have you learned describing in detail how this illustrates the Central Limit Theorem. Be sure to include a copy of *every* plot of sample means you evaluated.
2. Investigate how the sample means enable you to discriminate between two samples. Select Two Populations and select one of the three populations. Set both σ's to 5, sample sizes to 25, and μ's equal to one another. Draw 30 samples. Then increase one of the μ's by 1, update your graphs, and draw 30 more samples. Evaluate the overlap in the populations, the sampling distributions, and the dot plots of sample means. Repeat, continuing to separate the μ's until you can identify from which population *each* sample mean came. At that point, how far apart are the two means? Is this result consistent with the Central Limit Theorem? Write a report on what you have you learned. Do you think you could generalize this finding to other populations, σ's, or sample sizes? Include a copy of *every* population and sampling distribution you evaluated in your report.
3. Do project 2 with the following changes. Set both σ's to 8, set sample sizes to 25, and set one μ to be five units greater than the other μ. After evaluating the overlap, halve both σ's. Continue this process of halving the σ's until you can identify from which population *each* sample mean came. What is your value for σ? Is this result consistent with the Central Limit Theorem? Write a report on what you have learned. Do you think you could generalize this finding to other populations, sample sizes, or other differences in the μ's? Include a copy of *every* population and sampling distribution you evaluated.

Team Learning Projects

Select one of the three projects listed below. In each case, produce a team project that is suitable for an oral presentation. Use presentation software or large poster boards to display your results. Graphs should be large enough for your audience to see. Each team member should be responsible for producing some of the graphs. Ask your instructor if a written report is also expected.

1. A team of three can investigate the importance of sample size in discriminating between two samples. In the list box select Two Populations. Each team member should select a different population type (normal, uniform, skewed). Set both σ's to 2, sample sizes to 2, and set one μ to 1.0 greater than the other μ. Draw 20 samples. Each time evaluate the overlap in each of the following: populations, sampling distributions, and plot of sample means. Double both n's. Update your graphs, draw 20 more samples and again evaluate the overlaps. Continue this process until you can identify from which population *each* sample mean came. What is your value for n? Is this consistent with the Central Limit Theorem? Do you think you could generalize this finding if you sampled other populations, σ's, or differences in the μ's? Include illustrations of *every* population and sampling distribution you evaluated.
2. A team of three can investigate whether the sample mean is a consistent estimator for the population mean. Each team member should select a different population. Select any μ and σ. Set the sample size to 4. Draw 20 samples. Write down the average of the sample means and the standard error of the sample means. Repeat the process using sample sizes of 8, 16, 32, 64, and 128. How does this show that the sample mean is a consistent estimator for μ? Prepare an oral report on what you have you learned. Do you think you could generalize this finding if you sampled other populations? If you used other μ's? If you used other σ's? Be sure to include illustrations of the sampling distribution for *each* of the seven sample sizes for each population.
3. A team of three can investigate whether the sample mean is an unbiased estimator for the population mean. Each team member should select a different population. Select any μ and σ. Set the sample size to 2. Draw five samples. Write down the average of the sample means and the standard error of the sample means. Draw another five samples and write down the average of the sample means and their standard error. Continue drawing more samples and writing down the average of the sample means and their standard error for 20, 40, 60, 80, and 100 samples. How does this show that the sample mean is an unbiased estimator for μ? Prepare an oral report on what you have you learned. Do you think you could generalize this finding if you sampled other populations? If you used other μ's? If you used other σ's? If you used other sample sizes? Be sure to include illustrations of the sampling distribution showing *each* of your seven different numbers of samples (5, 10, 20, 40, 60, 80 and 100) for each population.

Self-Evaluation Quiz

1. The term "Central" in the Central Limit Theorem has which meaning?
   a. Middle.
   b. Fundamental.
   c. Central tendency.
   d. Sample mean.
   e. All of the above.

2. The Central Limit Theorem
   a. specifies certain asymptotic properties of the sample mean.
   b. is the basis for most statistical estimation.
   c. stems from work by many mathematicians, including De Moivre (1733), Laplace (1810), and Lindberg (1922).
   d. was discovered by Fra Luca Paciolo in 1494.
   e. all except d.

3. The Central Limit Theorem says
   a. that sample items will always form a normal, bell-shaped histogram.
   b. that the variance of the sample mean is the same as the population variance.
   c. that the variance of the sample mean exceeds the population variance.
   d. that the sample mean is a biased estimator of the mean if the population is not normal.
   e. none of the above.

4. If a normal population has parameters $\mu = 40$ and $\sigma = 8$, then for a sample of size $n = 4$
   a. the standard error of the sample mean is approximately 4.
   b. the standard error of the sample mean is approximately 2.
   c. the standard error of the sample mean is approximately 8.
   d. the standard error of the sample mean is approximately 10.
   e. the standard error depends on the population's shape.

5. If sample A with n items and sample B with 2n items are taken from a population, then
   a. the mean of sample A has a smaller expected variance than the mean of sample B.
   b. the mean of sample A has a greater expected variance than the mean of sample B.
   c. the ratio of the sample variances will be 2 to 1.
   d. the relative variance of the sample means will depend on the population shape.
   e. the variance of either sample will follow a normal distribution.

6. If the standard error of the mean is 12 when the sample size is 4, then if we increase the sample size to 16 the standard error will be
   a. 24
   b. 12
   c. 8
   d. 6
   e. None of the above.

7. The standard error of the mean is
   a. the standard deviation of the sample items.
   b. a measure of bias in sampling.
   c. the sample deviation of the standard normal distribution.
   d. an unbiased estimator of the population variance.
   e. none of the above.

8. The Central Limit Theorem applies to
   a. normally distributed populations.
   b. right-skewed populations.
   c. all populations with finite mean and variance.
   d. platykurtic populations.
   e. uniform populations.

9. The distribution of the mean of a sample of n items is
   a. normal if the population is normal and σ is known.
   b. approximately normal if n exceeds 20 and the population is moderately skewed.
   c. approximately normal if n exceeds 10 and the population is uniform.
   d. not always approximately normal if n exceeds 20.
   e. all of the above.

10. Which statement is correct?
    a. The sample mean is always an unbiased estimator of the population mean.
    b. The sample mean has a smaller variance than the population if n is at least 2.
    c. The sample mean has approximately a normal distribution if n is large.
    d. All of the above are correct.
    e. None of the above is correct.

11. Which is *not* true of a sample mean based on n items from a skewed, bimodal population with mean μ and variance σ^2?
    a. It will follow a bimodal distribution if n is very small.
    b. It will follow a normal distribution if n is large.
    c. It will be biased because of skewness.
    d. It will follow a skewed distribution if n is small.
    e. It has expected value μ.

12. A consistent estimator
    a. collapses on the true parameter as the population variance increases.
    b. collapses on the true parameter as the sample size increases.
    c. consistently follows a normal distribution.
    d. does not vary randomly from the true parameter.
    e. is impossible to obtain from real sample data.

Glossary of Terms

Asymptotic Statistical properties that are evident in sufficiently large samples. Usually, this refers to a sample size that approaches infinity (denoted as $n \to \infty$).

Central Limit Theorem This well-known theorem says that if the sample size is large enough, sample means from any population (even a non-normal population) will follow a normal distribution with the same mean as the population but with a smaller variance. More formally, if $x_1, x_2, \ldots, x_n$ are identically distributed independent random variables from any population with finite mean μ and finite variance σ^2, then the sample mean $\overline{X} = (x_1 + x_2 + \ldots + x_n) / n$ will have a sampling distribution with mean μ and variance σ^2/n. Further, the sampling distribution of $\overline{X}$ is asymptotically normally distributed.

Consistent estimator One that is unbiased (or asymptotically unbiased) and whose standard error decreases as the sample size increases. For example, the Central Limit Theorem says that the sample mean is a consistent estimator of μ, because as the sample size increases the distribution of the sample mean is centered at μ and becomes narrower (as $n \to \infty$, the distribution collapses on μ). This definition of consistency is also known as mean squared error consistency.

Mean For a population or a sampling distribution, the mean is the expected value of X, denoted as μ. It may be thought of as the probability-weighted average of the X values and may be interpreted as the fulcrum (balancing point) of the distribution along the X-axis. For a sample, the mean is the sum of the sample items divided by the sample size. It is denoted $\overline{X}$.

Normal population Bell-shaped or Gaussian distribution. It has two parameters: the mean μ and the variance σ^2.

Population being sampled In the Central Limit Theorem, the population from which the sample is drawn. It can have any shape, but it must have a finite mean and standard deviation.

Population distribution Probability density function f(x) showing the probability associated with each x value. The area under the entire distribution is 1 (or the sum of the probabilities if x is a discrete variable).

Population mean See **Mean**.

Sampling distribution Theoretical distribution of an estimator derived from a sample, such as the sample mean. According to the Central Limit Theorem, regardless of the population being sampled, the sample mean's sampling distribution has a mean μ and standard deviation $\sigma/\sqrt{n}$. It is normally distributed if the sample size is large enough.

Skewed population A population is skewed right and has a long right tail if its mean exceeds its median (and conversely, if the population is skewed left).

Standard deviation Denoted σ, it is the square root of the population variance. It is a measure of dispersion about the mean. Larger σ values indicate greater dispersion.

Standard error Standard deviation of the sample mean, given by the Central Limit Theorem. The theoretical standard error of the sample mean is generally denoted $\sigma_{\bar{x}}$ and its formula is $\sigma/\sqrt{n}$ where σ is the population standard deviation and n is the sample size. If the standard error is estimated from a sample, it is generally denoted $s_{\bar{x}}$ and its formula is $s/\sqrt{n}$ where s is the sample standard deviation and n is the sample size.

Unbiased estimator An estimator is unbiased if its expected value equals the parameter it is estimating. The Central Limit Theorem says that the sample mean is an unbiased estimator of μ. That is, the average of a large number of sample means will equal μ.

Uniform population A distribution that assigns the same probability to each X value in the domain $a \le X \le b$. It is a rectangle with base $b - a$ (the range of X) and height $1/(b - a)$.

Variance In a population, the variance is the expected value of $(X - \mu)^2$ and is denoted σ^2. In a sample, the variance is the sum of the squared deviations about the sample mean divided by $n - 1$ and is denoted s^2.

Solutions to Self-Evaluation Quiz

1. b Read the Overview of Concepts.
2. e Read the Overview of Concepts.
3. e Read the Overview of Concepts.
4. a Do Exercise 1. Read the Overview and the Illustration of Concepts.
5. b Do Exercises 1 and 4. Read the Overview of Concepts.
6. d Do Exercises 1–4. Read the Overview of Concepts.
7. e Do Exercise 1. Read the Overview of Concepts.
8. c Do Exercises 11–15. Read the Overview of Concepts.
9. e Do Exercises 8, 13–15, 18, and 19. Read the Illustration of Concepts.
10. d Read the Overview of Concepts. Do Exercises 21–25.
11. c Do Exercises 11–19 and 21–25.
12. b Do Exercise 26.