

that any particular confidence interval you happen to get is

We also show this graphically in the file. (See Figure 8.5.) The small square in this graph is positioned at the known mean and never changes. The blue line represents a particular confidence interval. Press the F9 key to force a recalculation. The position of the blue line will change. About 95% of the time, the blue line will straddle the small square—the confidence interval will include the true mean—but about 1 time out of 20, it will not. This also illustrates the meaning of a “95% confidence interval.”

Problems

Solutions for problems whose numbers appear within a colored box can be found in the Student Solution Files.

Level A

4. A manufacturing company's quality control personnel have recorded the proportion of defective items for each of 500 monthly shipments of one of the computer components that the company produces. The data are in the file **P07_07.xlsx**. The quality control department manager does not have sufficient time to review all of these data. Rather, she would like to examine the proportions of defective items for a sample of these shipments.
- a. Generate a simple random sample of size 25.

- b. Using the sample generated in part a, calculate a 95% confidence interval for the mean proportion of defective items over all monthly shipments. Assume that the population consists of the proportion of defective items for each of the given 500 monthly shipments.

- c. Interpret the 95% confidence interval constructed in part b.

- d. Does the 95% confidence interval contain the actual population mean in this case? If not, explain why not. What proportion of many similarly constructed confidence intervals should include the true population mean?

5. The file **P08_05.xlsx** contains salary data on all NFL players in each of the years 2002 to 2009. Because this file contains all players for each of these years, you can calculate the *population mean* for each year if *population*

¹ Depending on the speed of your computer, it can take a few seconds to simulate 1000 samples of size 30 in this data table. Therefore, it is a good idea to set the recalculation mode to “automatic except tables.” (You can find this option under the Calculation Options dropdown menu on the Formulas ribbon.) That way, the data table recalculates only if you explicitly tell it to (by pressing the F9 key).

Level B

sample of size 50 from the population. Based on this random sample, calculate a 95% confidence interval for the mean NFL total salary in 2009. Does it contain the population mean? Repeat this procedure several times until you find a random sample where the population mean is *not* included in the confidence interval.

6. The file **P08_06.xlsx** contains data on repetitive task times for each of two workers. John has been doing this task for months, whereas Fred has just started. Each time listed is the time (in seconds) to perform a routine task on an assembly line. The times shown are in chronological order.

a. Calculate a 95% confidence interval for the mean time it takes John to perform the task. Do the same for

Fred.

Fred.

b. Do you believe both of the confidence intervals in part a are valid and/or useful? Why or why not? Which of the two workers would you rather have, assuming that task time is the only issue?

7. The manager of a local fast-food restaurant is interested in improving the service provided to customers who use the restaurant's drive-up window. As a first step in this process, the manager asks an assistant to record the time (in seconds) it takes to serve a large number of customers at the final window in the facility's drive-up system.

8. Continuing Problem 5, find confidence intervals for each of the 6 players. For each of these same players, find a confidence interval for the mean number of runs scored in the confidence level that you choose. Do the confidence intervals include the mean number of runs scored in the year? Is this the same for all players? Do these confidence intervals include the mean number of runs scored in the year? What does this mean? Why or why not?

9. The file **Confidence** contains 100 observations randomly generated to follow a normal distribution. Suppose instead that each observation was generated from a non-normally distributed population. Write a brief explanation for a brief explanation of how you would know. Unlike a normal distribution, a non-normal distribution is very skewed to the right. The data can be generated using the `rnorm()` function. Rerun the simulation using the `rexp()` function to generate exponential distributed random variables. Are the data still *valid*? That is, are the confidence intervals approximately equal to the nominal level of a few times to check. 10. Answer the question: Is the population a normal distribution?

10. Answer the question. The population is a locally supported

17.

a. Calculate a 90% confidence interval for the proportion of all the university's employees who favor plan A. b. The file also includes the classification of each employee (administrative staff, support staff, or faculty). Calculate a separate 90% confidence interval for each of these groups for the proportion who favor plan A. How do these confidence intervals compare to one another? How do their lengths compare to the confidence interval in part a? Is this what you would expect? Explain.

A market research consultant hired by a leading soft-drink company wants to determine the proportion of consumers who favor its low-calorie brand over the leading low-calorie competitor in a particular geographic region. A random sample of 250 consumers from the market under investigation is provided in the file **P08_17.xlsx**.

a. Calculate a 90% confidence interval for the proportion of all consumers in this market who prefer the company's brand.

b. The file contains the gender and age group for each customer in the sample. Calculate a separate 90% confidence for each gender for the proportion who prefer the company's brand. Then do the same for each age group. Explain briefly how these confidence intervals compare to each other and to the confidence interval in part a.

Level A

- 20.** The director of a university's career development center is interested in comparing the starting annual salaries of male and female students who recently graduated from the university and commenced full-time employment. The director has formed pairs of male and female graduates with the same major and similar grade-point averages. Specifically, she has collected a random sample of 50 such pairs and has recorded the starting annual salary of each person. These data are provided in the file **P08_20.xlsx**. Calculate a 95% confidence interval for the mean difference between similar male and female graduates of this university. Interpret your result.
- 21.** A real estate agent has collected a random sample of 75 houses that were recently sold in a suburban community. She is particularly interested in comparing the appraised value and recent selling price of the houses in this particular market. The data are provided in the file **P08_21.xlsx**. Using this sample data, calculate a 95% confidence interval for the mean difference between the appraised values and selling prices of the houses sold in this suburban community. Interpret the confidence interval for the real estate agent.
- 22.** *The Wall Street Journal CEO Compensation Study* analyzed CEO pay from many U.S. companies with fiscal year 2008 revenue of at least \$5 billion that filed their proxy statements between October 2008 and March 2009. The data are in the file **P02_30.xlsx**.
- a.** Create a new column, Total, that is the sum of columns D and E.
- b.** After combining Telecommunications and Technology into a single company type, there are nine company types. For each of these, calculate a 95% confidence interval for the difference between the mean of Total for

Problems

Solutions for problems whose numbers appear within a colored box can be found in the Student Solution Files.

Level A

25. A company that advertises on the Web wants to know which search engine its customers prefer as their primary search engine: Google or Bing. Specifically, the company wants to know whether the preference depends on the browser being used. The file **P08_25.xlsx** contains counts of 800 customers' favorite search engine, broken down by the browser used.

a. Calculate a 95% confidence interval for the difference between two proportions: the proportion of Internet Explorer users whose favorite search engine is Google and the similar proportion of Firefox users.

b. Repeat part a, replacing Google with Bing.

c. Interpret the results in parts a and b. Do the search engine preferences seem to depend on the browser used?

26. A market research consultant hired by a leading soft-drink company is interested in estimating the difference between the proportions of female and male consumers who favor the company's low-calorie brand over the leading competitor's low-calorie brand in a particular geographical region. A random sample of 250 consumers from the market is selected. The file **P08_17.xlsx** contains the data.