Solution

A 95% confidence interval has width $\pm 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$. The widest the confidence interval could be, for a sample of size n, is $\pm 1.96\sqrt{(0.5)(1-0.5)/(n+4)}$, or $\pm 0.98/\sqrt{n+4}$. Solving the equation $0.98/\sqrt{n+4} = 0.08$ for n, we obtain $n \approx 147$. Note that this estimate is somewhat larger than the one obtained in Example 5.13.

The Traditional Method

The method we have described was developed quite recently (although it was created by simplifying a much older method). Many people still use a more traditional method. The traditional method uses the actual sample size n in place of $\tilde{n}$, and the actual sample proportion $\widehat{p}$ in place of $\tilde{p}$. Although this method is still widely used, it fails to achieve its stated coverage probability even for some fairly large values of n. This means that $100(1-\alpha)\%$ confidence intervals computed by the traditional method will cover the true proportion less than $100(1-\alpha)\%$ of the time. The traditional method cannot be used for small samples at all; one rule of thumb regarding the sample size is that both $n\widehat{p}$ (the number of successes) and $n(1-\widehat{p})$ (the number of failures) should be greater than 10.

Since the traditional method is still widely used, we summarize it in the following box. For very large sample sizes, the results of the traditional method are almost identical to those of the modern method. For small or moderately large sample sizes, the modern approach is better.

Summary

The Traditional Method for Computing Confidence Intervals for a Proportion (widely used but not recommended)

Let $\widehat{p}$ be the proportion of successes in a *large* number n of independent Bernoulli trials with success probability p. Then the traditional level $100(1-\alpha)\%$ confidence interval for p is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \tag{5.8}$$

The method cannot be used unless the sample contains at least 10 successes and 10 failures.

Exercises for Section 5.2

1. In a simple random sample of 70 automobiles registered in a certain state, 28 of them were found to have emission levels that exceed a state standard.

a. What proportion of the automobiles in the sample had emission levels that exceed the standard?

b. Find a 95% confidence interval for the proportion of automobiles in the state whose emission levels exceed the standard.

c. Find a 98% confidence interval for the proportion of automobiles whose emission levels exceed the standard.

d. How many automobiles must be sampled to specify the proportion that exceed the standard to within ± 0.10 with 95% confidence?

e. How many automobiles must be sampled to specify the proportion that exceed the standard to within ± 0.10 with 98% confidence?

f. Someone claims that less than half of the automobiles in the state exceed the standard. With what level of confidence can this statement be made?

2. During a recent drought, a water utility in a certain town sampled 100 residential water bills and found that 73 of the residences had reduced their water consumption over that of the previous year.

a. Find a 95% confidence interval for the proportion of residences that reduced their water consumption.

b. Find a 99% confidence interval for the proportion of residences that reduced their water consumption.

c. Find the sample size needed for a 95% confidence interval to specify the proportion to within ± 0.05.

d. Find the sample size needed for a 99% confidence interval to specify the proportion to within ± 0.05.

e. Someone claims that more than 70% of residences reduced their water consumption. With what level of confidence can this statement be made?

f. If 95% confidence intervals are computed for 200 towns, what is the probability that more than 192 of the confidence intervals cover the true proportions?

3. A soft-drink manufacturer purchases aluminum cans from an outside vendor. A random sample of 70 cans is selected from a large shipment, and each is tested for strength by applying an increasing load to the side of the can until it punctures. Of the 70 cans, 52 meet the specification for puncture resistance.

a. Find a 95% confidence interval for the proportion of cans in the shipment that meet the specification.

b. Find a 90% confidence interval for the proportion of cans in the shipment that meet the specification.

c. Find the sample size needed for a 95% confidence interval to specify the proportion to within ± 0.05.

d. Find the sample size needed for a 90% confidence interval to specify the proportion to within ± 0.05.

e. If a 90% confidence interval is computed each day for 300 days, what is the probability that more than 280 of the confidence intervals cover the true proportions?

4. The article "HIV-positive Smokers Considering Quitting: Differences by Race/Ethnicity" (E. Lloyd-Richardson, C. Stanton, et al., *Am J Health Behav*, 2008:3–15) surveyed 444 HIV-positive smokers. Of these, 170 reported that they had used a nicotine patch. Consider this to be a simple random sample.

a. Find a 95% confidence interval for the proportion of HIV-positive smokers who have used a nicotine patch.

b. Find a 99% confidence interval for the proportion of HIV-positive smokers who have used a nicotine patch.

c. Someone claims that the proportion is less than 0.40. With what level of confidence can this statement be made?

d. Find the sample size needed for a 95% confidence interval to specify the proportion to within ± 0.03.

e. Find the sample size needed for a 99% confidence interval to specify the proportion to within ± 0.03.

5. The article "The Functional Outcomes of Total Knee Arthroplasty" (R. Kane, K. Saleh, et al., *Journal of Bone and Joint Surgery,* 2005:1719–1724) reports that out of 10,501 surgeries, 859 resulted in complications within six months of surgery.

a. Find a 95% confidence interval for the proportion of surgeries that result in complications within six months.

b. Find a 99% confidence interval for the proportion of surgeries that result in complications within six months.

c. A surgeon claims that the rate of complications is less than 8.5%. With what level of confidence can this claim be made?

6. Refer to Exercise 1. Find a 95% lower confidence bound for the proportion of automobiles whose emissions exceed the standard.

7. Refer to Exercise 2. Find a 98% upper confidence bound for the proportion of residences that reduced their water consumption.

8. Refer to Exercise 4. Find a 99% lower confidence bound for the proportion of HIV-positive smokers who have used a nicotine patch.

9. A random sample of 400 electronic components manufactured by a certain process are tested, and 30 are found to be defective.

a. Let p represent the proportion of components manufactured by this process that are defective. Find a 95% confidence interval for p.

b. How many components must be sampled so that the 95% confidence interval will specify the proportion defective to within ± 0.02?

c. (Hard) The company ships the components in lots of 200. Lots containing more than 20 defective components may be returned. Find a 95% confidence interval for the proportion of lots that will be returned.

10. Refer to Exercise 9. A device will be manufactured in which two of the components in Exercise 9 will be connected in series. The components function independently, and the device will function only if both components function. Let q be the probability that a device functions. Find a 95% confidence interval for q. (*Hint:* Express q in terms of p, and then use the result of Exercise 9a.)

11. When the light turns yellow, should you stop or go through it? The article "Evaluation of Driver Behavior in Type II Dilemma Zones at High-Speed Signalized Intersections" (D. Hurwitz, M. Knodler, and B. Nyquist, *Journal of Transportation Engineering* 2011:277–286) defines the "indecision zone" as the period when a vehicle is between 2.5 and 5.5 seconds away from an intersection. It presents observations of 710 vehicles passing through various intersections in Vermont for which the light turned yellow in the indecision zone. Of the 710 vehicles, 89 ran a red light.

a. Find a 90% confidence interval for the proportion of vehicles that will run the red light when encountering a yellow light in the indecision zone.

b. Find a 95% confidence interval for the proportion of vehicles that will run the red light when encountering a yellow light in the indecision zone.

c. Find a 99% confidence interval for the proportion of vehicles that will run the red light when encountering a yellow light in the indecision zone.

12. In a random sample of 150 customers of a high-speed internet provider, 63 said that their service had been interrupted one or more times in the past month.

a. Find a 95% confidence interval for the proportion of customers whose service was interrupted one or more times in the past month.

b. Find a 99% confidence interval for the proportion of customers whose service was interrupted one or more times in the past month.

c. Find the sample size needed for a 95% confidence interval to specify the proportion to within ± 0.05.

d. Find the sample size needed for a 99% confidence interval to specify the proportion to within ± 0.05.

13. A sociologist is interested in surveying workers in computer-related jobs to estimate the proportion of such workers who have changed jobs within the past year.

a. In the absence of preliminary data, how large a sample must be taken to ensure that a 95% confidence interval will specify the proportion to within ± 0.05?

b. In a sample of 100 workers, 20 of them had changed jobs within the past year. Find a 95% confidence interval for the proportion of workers who have changed jobs within the past year.

c. Based on the data in part (b), estimate the sample size needed so that the 95% confidence interval will specify the proportion to within ± 0.05.

14. Stainless steels can be susceptible to stress corrosion cracking under certain conditions. A materials engineer is interested in determining the proportion of steel alloy failures that are due to stress corrosion cracking.

a. In the absence of preliminary data, how large a sample must be taken so as to be sure that a 98% confidence interval will specify the proportion to within ± 0.05?

b. In a sample of 200 failures, 30 of them were caused by stress corrosion cracking. Find a 98% confidence interval for the proportion of failures caused by stress corrosion cracking.

c. Based on the data in part (b), estimate the sample size needed so that the 98% confidence interval will specify the proportion to within ± 0.05.

15. The article "A Music Key Detection Method Based on Pitch Class Distribution Theory" (J. Sun, H. Li, and L. Ma, *International Journal of Knowledge-based and Intelligent Engineering Systems*, 2011:165–175) describes a method of analyzing digital music files to determine the key in which the music is written. In a sample of 335 classical music selections, the key was identified correctly in 293 of them.

a. Find a 90% confidence interval for the proportion of pieces for which the key will be correctly identified.

b. How many music pieces should be sampled to specify the proportion to within ± 0.025 with 90% confidence?

c. Another method of key detection is to be tested. At this point, there is no estimate of the proportion of the time this method will be identified correctly. Find a conservative estimate of the sample size needed so that the proportion will be specified to within ± 0.03 with 90% confidence.

16. A stock market analyst notices that in a certain year, the price of IBM stock increased on 131 out of 252 trading days. Can these data be used to find a 95% confidence interval for the proportion of days that IBM stock increases? Explain.

5.3 Small-Sample Confidence Intervals for a Population Mean

The methods described in Section 5.1 for computing confidence intervals for a population mean require that the sample size be large. When the sample size is small, there are no good general methods for finding confidence intervals. However, when the population is approximately normal, a probability distribution called the Student's t distribution can be used to compute confidence intervals for a population mean. In this section, we describe this distribution and show how to use it.

The Student's *t* Distribution

If $\overline{X}$ is the mean of a large sample of size n from a population with mean μ and variance σ^2, then the Central Limit Theorem specifies that $\overline{X} \sim N(\mu, \sigma^2/n)$. The quantity $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ then has a normal distribution with mean 0 and variance 1. In addition, the sample standard deviation s will almost certainly be close to the population standard deviation σ. For this reason the quantity $(\overline{X} - \mu)/(s/\sqrt{n})$ is approximately normal with mean 0 and variance 1, so we can look up probabilities pertaining to this quantity in the standard normal table (z table). This enables us to compute confidence intervals of various levels for the population mean μ.

What can we do if $\overline{X}$ is the mean of a *small* sample? If the sample size is small, s may not be close to σ, and $\overline{X}$ may not be approximately normal. If we know nothing about the population from which the small sample was drawn, there are no easy methods for computing confidence intervals. However, if the population is approximately normal, $\overline{X}$ will be approximately normal even when the sample size is small. It turns out that we can still use the quantity $(\overline{X} - \mu)/(s/\sqrt{n})$, but since s is not necessarily close to σ, this quantity will not have a normal distribution. Instead, it has the **Student's *t* distribution** with $n - 1$ degrees of freedom, which we denote t_{n-1}. The number of degrees of freedom for the t distribution is one less than the sample size.

The Student's t distribution was discovered in 1908 by William Sealy Gossett, a statistician who worked for the Guinness Brewing Company in Dublin, Ireland. The management at Guinness considered the discovery to be proprietary information, and forbade Gossett to publish it. He published it anyway, using the pseudonym "Student." Gossett had done this before; see Section 4.3.