SUMMARY

The approximate **level *C* confidence interval** for the difference in two proportions $p_1 - p_2$ is

$$(\tilde{p}_1 - \tilde{p}_2) \pm z^* \mathrm{SE}_{\tilde{D}}$$

where the **Wilson estimates** of the population proportions are

$$\tilde{p}_1 = \frac{X_1 + 1}{n_1 + 2} \quad \text{and} \quad \tilde{p}_2 = \frac{X_2 + 1}{n_2 + 2}$$

and the **standard error of the difference** is

$$\mathrm{SE}_{\tilde{D}} = \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2 + 2}}$$

Significance tests of H_0: $p_1 = p_2$ use the ***z* statistic**

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\mathrm{SE}_{Dp}}$$

with *P*-values from the $N(0, 1)$ distribution. In this statistic,

$$\mathrm{SE}_{Dp} = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and $\hat{p}$ is the **pooled estimate** of the common value of p_1 and p_2:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Relative risk is the ratio of two sample proportions:

$$\mathrm{RR} = \frac{\hat{p}_1}{\hat{p}_2}$$

Confidence intervals for relative risk are often used to summarize the comparison of two proportions.

SECTION 8.2 EXERCISES

8.31 Different kinds of companies compensate their key employees in different ways. Established companies may pay higher salaries, while new companies may offer stock options that will be valuable if the company succeeds. Do

high-tech companies tend to offer stock options more often than other companies? One study looked at a random sample of 200 companies. Of these, 91 were listed in the *Directory of Public High Technology Corporati* and 109 were not listed. Treat these two groups as SRSs of high-tech and non-high-tech companies. Seventy-three of the high-tech companies and 75 of the non-high-tech companies offered incentive stock options to key employees.[11] Give a 95% confidence interval for the difference in the proportions of the two types of companies that offer stock options.

8.32 Refer to the previous exercise. Compare the two groups of companies wi significance test. Summarize your analysis and conclusions.

8.33 A study was designed to find reasons why patients leave a health maintenance organization (HMO).[12] Patients were classified as to wheth or not they had filed a complaint with the HMO. We want to compare the proportion of complainers who leave the HMO with the proportion of those who do not file complaints. In the year of the study, 639 patient filed complaints, and 54 of these patients left the HMO voluntarily. For comparison, the HMO chose an SRS of 743 patients who had not filed complaints. Twenty-two of these patients left voluntarily. Give an estima of the difference in the two proportions with a 95% confidence interval.

8.34 In the previous exercise you examined data from a study designed to fin reasons why patients leave an HMO. There you compared the proportio of complainers who leave the HMO with the proportion of noncomplain who leave. In the year of the study, 639 patients filed complaints and 54 these patients left the HMO voluntarily. For comparison, the HMO chos an SRS of 743 patients who had not filed complaints. Twenty-two of tho patients left voluntarily. We expect a higher proportion of complainers t leave. Do the data support this expectation? State hypotheses, find the te statistic and its P-value, and state your conclusion.

8.35 To what extent do syntax textbooks, which analyze the structure of sentences, illustrate gender bias? A study of this question sampled sente from 10 texts.[13] One part of the study examined the use of the words "gi "boy," "man," and "woman." We will call the first two words *juvenile* and last two *adult*. Is the proportion of female references that are juvenile (g equal to the proportion of male references that are juvenile (boy)? Here data from one of the texts:

Gender	n	X(juvenile)
Female	60	48
Male	132	52

(a) Find the proportion of juvenile references for females and its stand error. Do the same for the males.

(b) Give a 90% confidence interval for the difference and briefly summa what the data show.

8.36 In Exercise 8.6 (page 582) we examined the percent of fatally injured bicyclists tested for alcohol who tested positive. Here we examine the same data with respect to gender.

Gender	n	X(tested positive)
Female	191	27
Male	1520	515

(a) Summarize the data by giving the Wilson estimates of the two population proportions and a 95% confidence interval for their difference.

(b) The standard error $SE_{\tilde{D}}$ contains a contribution from each sample, $\tilde{p}_1(1-\tilde{p}_1)/(n_1+2)$ and $\tilde{p}_2(1-\tilde{p}_2)/(n_2+2)$. Which of these contributes the larger amount to the standard error of the difference? Explain why.

8.37 Is lying about credentials by job applicants changing? In Exercise 8.4 (page 582) we looked at the proportion of applicants who lied about having a degree in a six-month period. To see if there is a change over time, we can compare that period with the following six months. Here are the data:

Period	n	X(lied)
1	84	15
2	106	21

Use a 95% confidence interval to address the question of interest.

8.38 Exercise 8.35 addresses a question about gender bias with a confidence interval. Set up the problem as a significance test. Carry out the test and summarize the results.

8.39 The proportions of female and male fatally injured bicyclists were compared with a confidence interval in Exercise 8.36. Examine the same data with a test of significance.

8.40 Data on the proportion of applicants who lied about having a degree in two consecutive six-month periods are given in Exercise 8.37. Formulate appropriate null and alternative hypotheses that can be addressed with these data, carry out the significance test, and summarize the results.

8.41 In the Christmas tree survey introduced in Example 8.6 (page 580), respondents who had a tree during the holiday season were asked whether the tree was natural or artificial. Respondents were also asked if they lived in an urban area or in a rural area. Of the 421 households displaying a Christmas tree, 160 lived in rural areas and 261 were urban residents. The

tree growers want to know if there is a difference in preference for natu trees versus artificial trees between urban and rural households. Here a the data:

Population	n	X(natural)
1 (rural)	160	64
2 (urban)	261	89

(a) Give the null and alternative hypotheses that are appropriate for th problem assuming that we have no prior information suggesting th one population would have a higher preference than the other.

(b) Test the null hypothesis. Give the test statistic and the P-value, and summarize the results.

(c) Give a 95% confidence interval for the difference in proportions.

8.42 In the 2000 regular baseball season, the World Series Champion New Y Yankees played 80 games at home and 81 games away. They won 44 of home games and 43 of the games played away. We can consider these g as samples from potentially large populations of games played at home away. How much advantage does the Yankee home field provide?

(a) Find the Wilson estimate of proportion of wins for all home games the same for away games.

(b) Find the standard error needed to compute a confidence interval fo difference in the proportions.

(c) Compute a 90% confidence interval for the difference between the probability that the Yankees win at home and the probability that t win when on the road. Are you convinced that the Yankees were m likely to win at home in 2000?

8.43 Refer to the New York Yankees baseball data in the previous exercise.

(a) Combining all of the games played, what proportion did the Yankee win?

(b) Find the standard error needed for testing that the probability of winning is the same at home and away.

(c) Most people think that it is easier to win at home than away. Formu null and alternative hypotheses to examine this idea.

(d) Compute the z statistic and its P-value. What conclusion do you dra

8.44 In the 2000 World Series the New York Yankees played the New York Me The previous two exercises examine the Yankees' home and away victori During the regular season the Mets won 55 of the 84 home games that th played and 39 of the 81 games that they played away. Perform the same analyses for the Mets and write a short summary comparing these result with those you found for the Yankees.

8.45 The state agriculture department asked random samples of Indiana farmers in each county whether they favored a mandatory corn checkoff program to pay for corn product marketing and research. In Tippecanoe County, 263 farmers were in favor of the program and 252 were not. In neighboring Benton County, 260 were in favor and 377 were not.

(a) Find the proportions of farmers in favor of the program in each of the two counties.

(b) Find the standard error needed to compute a confidence interval for the difference in the proportions.

(c) Compute a 95% confidence interval for the difference between the proportions of farmers favoring the program in Tippecanoe County and in Benton County. Do you think opinions differed in the two counties?

8.46 Return to the survey of farmers described in the previous exercise.

(a) Combine the two samples and find the overall proportion of farmers who favor the corn checkoff program.

(b) Find the standard error needed for testing that the population proportions of farmers favoring the program are the same in the two counties.

(c) Formulate null and alternative hypotheses for comparing the two counties.

(d) Compute the z statistic and its P-value. What conclusion do you draw?

8.47 A major court case on the health effects of drinking contaminated water took place in the town of Woburn, Massachusetts. A town well in Woburn was contaminated by industrial chemicals. During the period that residents drank water from this well, there were 16 birth defects among 414 births. In years when the contaminated well was shut off and water was supplied from other wells, there were 3 birth defects among 228 births. The plaintiffs suing the firm responsible for the contamination claimed that these data show that the rate of birth defects was higher when the contaminated well was in use.[14] How statistically significant is the evidence? What assumptions does your analysis require? Do these assumptions seem reasonable in this case?

8.48 A study of chromosome abnormalities and criminality examined data on 4124 Danish males born in Copenhagen.[15] The study used the penal registers maintained in the offices of the local police chiefs and classified each man as having a criminal record or not. Each was also classified as having the normal male XY chromosome pair or one of the abnormalities XYY or XXY. Of the 4096 men with normal chromosomes, 381 had criminal records, while 8 of the 28 men with chromosome abnormalities had criminal records. Some experts believe that chromosome abnormalities are associated with increased criminality. Do these data lend support to this belief? Report your analysis and draw a conclusion.

8.49 A university financial aid office polled an SRS of undergraduate stude to study their summer employment. Not all students were employed t previous summer. Here are the results for men and women:

	Men	Women
Employed	728	603
Not employed	89	149
Total	817	752

(a) Is there evidence that the proportion of male students employed the summer differs from the proportion of female students who w employed? State H_0 and H_a, compute the test statistic, and give it *P*-value.

(b) Give a 95% confidence interval for the difference between the proportions of male and female students who were employed dur the summer. Does the difference seem practically important to y

8.50 A clinical trial examined the effectiveness of aspirin in the treatment cerebral ischemia (stroke). Patients were randomized into treatment and control groups. The study was double-blind in the sense that neit the patients nor the physicians who evaluated the patients knew whic patients received aspirin and which the placebo tablet.[16] After six mon of treatment, the attending physicians evaluated each patient's progres as either favorable or unfavorable. Of the 78 patients in the aspirin gr 63 had favorable outcomes; 43 of the 77 control patients had favorabl outcomes.

(a) Compute the sample proportions of patients having favorable out in the two groups.

(b) Give a 90% confidence interval for the difference between the fav proportions in the treatment and control groups.

(c) The physicians conducting the study had concluded from previous research that aspirin was likely to increase the chance of a favora outcome. Carry out a significance test to confirm this conclusion. S hypotheses, find the *P*-value, and write a summary of your results.

8.51 The pesticide diazinon is in common use to treat infestations of the G cockroach, *Blattella germanica*. A study investigated the persistence of this pesticide on various types of surfaces.[17] Researchers applied a 0.5 emulsion of diazinon to glass and plasterboard. After 14 days, they pla 18 cockroaches on each surface and recorded the number that died wi 48 hours. On glass, 9 cockroaches died, while on plasterboard, 13 died.

(a) Calculate the mortality rates (sample proportion that died) for the surfaces.

(b) Find a 95% confidence interval for the difference in the two popul proportions.

(c) Chemical analysis of the residues of diazinon suggests that it may persist longer on plasterboard than on glass because it binds to the paper covering on the plasterboard. The researchers therefore expected the mortality rate to be greater on plasterboard than on glass. Conduct a significance test to assess the evidence that this is true.

8.52 Refer to the study of undergraduate student summer employment described in Exercise 8.49. Similar results from a smaller number of students may not have the same statistical significance. Specifically, suppose that 73 of 82 men surveyed were employed and 60 of 75 women surveyed were employed. The sample proportions are essentially the same as in the earlier exercise.

(a) Compute the z statistic for these data and report the P-value. What do you conclude?

(b) Compare the results of this significance test with your results in Exercise 8.49. What do you observe about the effect of the sample size on the results of these significance tests?

8.53 Suppose that the experiment of Exercise 8.51 placed more cockroaches on each surface and observed similar mortality rates. Specifically, suppose that 36 cockroaches were placed on each surface and that 26 died on the plasterboard, while 18 died on the glass.

(a) Compute the z statistic for these data and report its P-value. What do you conclude?

(b) Compare the results of this significance test with those you gave in Exercise 8.51. What do you observe about the effect of the sample size on the results of these significance tests?

CHAPTER 8 EXERCISES

8.54 "The nature of work is changing at whirlwind speed. Perhaps now more than ever before, job stress poses a threat to the health of workers and, in turn, to the health of organizations."[18] So says the National Institute for Occupational Safety and Health. Employers are concerned about the effect of stress on their employees. Stress can lower morale and efficiency and increase medical costs. A large survey of restaurant employees found that 75% reported that work stress had a negative impact on their personal lives.[19] The human resources manager of a chain of restaurants is concerned that work stress may be affecting the chain's employees. She asks a random sample of 100 employees to respond Yes or No to the question "Does work stress have a negative impact on your personal life?" Of these, 68 say "Yes." Give a 95% confidence interval for the proportion of employees who work for this chain of restaurants who believe that work stress has a negative impact on their personal lives.

8.55 Refer to the previous exercise. Is there evidence to conclude that the proportion for this chain of restaurants differs from the value given for the