8-1 Overview

Chapter 6 introduced an important activity of inferential statistics: Sample data were used to construct confidence interval *estimates* of population parameters. Chapter 7 introduced a second important activity of inferential statistics: Sample data were used to *test hypotheses* about population parameters. In Chapters 6 and 7, all examples and exercises involved the use of *one* sample to form an inference about *one* population. In reality, there are many important and meaningful situations in which it becomes necessary to compare *two* sets of sample data. The following are examples typical of those found in this chapter, which presents methods for using sample data from two populations so that inferences can be made about those populations.

- When testing the effectiveness of the Salk vaccine in preventing paralytic polio, determine whether the treatment group had a lower incidence of polio than the group given a placebo.
- When testing the effectiveness of a drug designed to lower cholesterol, determine whether the treatment group had more substantial reductions than the group given a placebo.
- When investigating the accuracy of heights reported by people, determine whether there is a significant difference between the heights they report and the actual measured heights.

Chapters 6 and 7 included methods that were applied to proportions, means, and measures of variation (standard deviation and variance), and this chapter will address those same parameters. This chapter extends the same methods introduced in Chapters 6 and 7 to situations involving comparisons of two samples instead of only one. Also, Section 8-5 presents a method for finding confidence interval estimates of odds ratios. (Odds ratios were first introduced in Section 3-6.)

8-2 Inferences About Two Proportions

There are many real and important situations in which it is necessary to use sample data to compare two population proportions. Media reports and articles in professional journals include an abundance of situations involving a comparison of two population proportions. Although this section is based on proportions, we can deal with probabilities or we can deal with percentages by using the corresponding decimal equivalents.

When testing a hypothesis made about two population proportions or when constructing a confidence interval estimate of the difference between two population proportions, the methods presented in this section have the following requirements. (In addition to the methods included in this section, there are other methods for making inferences about two proportions, and they have different requirements.)

Requirements for Inferences About Two Proportions

1. We have proportions from two simple random samples that are *independent*, which means that the sample values selected from one population are not related to or somehow paired or matched with the sample values selected from the other population. (For example, in analyzing differences between headache rates among men and headache rates among women, we should not use a sample in which all of the men are husbands of the corresponding women. Such a sample would involve matched sample values instead of independent samples.)
2. For each of the two samples, the number of successes is at least five, and the number of failures is at least five.

Notation for Two Proportions

For population 1 we let

p_1 = *population* proportion

n_1 = size of the sample taken from population 1

x_1 = number of successes in the sample taken from population 1

$\hat{p}_1 = \frac{x_1}{n_1}$ (the *sample* proportion)

$\hat{q}_1 = 1 - \hat{p}_1$

The corresponding meanings are attached to $p_2, n_2, x_2, \hat{p}_2$, and $\hat{q}_2$, which come from population 2. Note that we use p_1 to denote the proportion of the entire *population*, but we use $\hat{p}_1$ to denote the proportion found from the *sample.*

Finding the Numbers of Successes x_1 and x_2: The calculations for hypothesis tests and confidence intervals require that we have specific values for x_1, n_1, x_2, and n_2. Sometimes the available sample data include those specific numbers, but sometimes it is necessary to calculate the values of x_1 and x_2.

For example, consider the statement that "when 734 men were treated with Viagra, 16% of them experienced headaches." From that statement we can see that $n_1 = 734$ and $\hat{p}_1 = 0.16$, but the actual number of successes x_1 is not given. However, from $\hat{p}_1 = x_1/n_1$, we know that

$$x_1 = n_1 \cdot \hat{p}_1$$

so that $x_1 = 734 \cdot 0.16 = 117.44$. But you cannot have 117.44 men who experienced headaches, because each man either experiences a headache or does not, and the number of successes x_1 must therefore be a whole number. We can round 117.44 to 117. We can now use $x_1 = 117$ in the calculations that require its value. It's really quite simple: 16% of 734 means 0.16×734, which results in 117.44, which we round to 117.

Hypothesis Tests

In Section 7-2 we discussed tests of hypotheses made about a single population proportion. We will now consider tests of hypotheses made about two population proportions, but *we will be testing only claims that* $p_1 = p_2$, and we will use the following pooled (or combined) estimate of the value that p_1 and p_2 have in common. (For claims that the difference between p_1 and p_2 is equal to a nonzero constant, see Exercise 23 of this section.) You can see from the form of the pooled estimate $\bar{p}$ that it basically combines the two different samples into one big sample.

Pooled Estimate of p_1 and p_2

The **pooled estimate of p_1 and p_2** is denoted by $\bar{p}$ and is given by

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

We denote the complement of $\bar{p}$ by $\bar{q}$, so $\bar{q} = 1 - \bar{p}$.

Test Statistic for Two Proportions (with H_0: $p_1 = p_2$)

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\bar{p}\,\bar{q}}{n_1} + \dfrac{\bar{p}\,\bar{q}}{n_2}}}$$

where $p_1 - p_2 = 0$ (assumed in the null hypothesis)

$$\hat{p}_1 = \frac{x_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\bar{q} = 1 - \bar{p}$$

***P*-value:** Use Table A-2. (Use the computed value of the test statistic z and find the *P*-value by following the procedure summarized in Figure 7-6.)

Critical values: Use Table A-2. (Based on the significance level α, find critical values by using the procedures introduced in Section 7-2.)

The following example will help clarify the roles of x_1, n_1, $\hat{p}_1$, $\bar{p}$, and so on. In particular, you should recognize that under the assumption of equal proportions, the best estimate of the common proportion is obtained by pooling both samples into one big sample, so that $\bar{p}$ becomes a more obvious estimate of the common population proportion.

EXAMPLE Testing Effectiveness of a Vaccine In the Chapter Problem, we noted that a *USA Today* article reported experimental results from a nasal spray vaccine for children. Among 1070 children given the vaccine, 14 developed a flu. Among 532 children given a placebo, 95 developed a flu. The sample data are summarized in Table 8-1. Use a 0.05 significance level to test the claim that the proportion of flu cases among the vaccinated children is less than the proportion of flu cases among children given a placebo.

SOLUTION For notation purposes, we stipulate that Sample 1 is the treatment (vaccinated) group, and Sample 2 is the placebo group. We can summarize

the sample data as follows. (We use extra decimal places for the sample proportions $\hat{p}_1$ and $\hat{p}_2$ because those values will be used in subsequent calculations.)

Vaccinated Children	Children Given Placebo
$n_1 = 1070$	$n_2 = 532$
$x_1 = 14$	$x_2 = 95$
$\hat{p}_1 = \dfrac{x_1}{n_1} = \dfrac{14}{1070} = 0.013084$	$\hat{p}_2 = \dfrac{x_2}{n_2} = \dfrac{95}{532} = 0.178571$

REQUIREMENT ✓ We should first verify that the necessary requirements (listed earlier in the section) are satisfied. Given the design of the experiment, it is reasonable to assume that both samples are simple random samples and they are independent. Also, each of the two samples has at least five successes and at least five failures. (The first sample has 14 successes and 1056 failures. The second sample has 95 successes and 437 failures.) The check of requirements has been successfully completed and we can now proceed to conduct the formal hypothesis test. ✓

We will now use the *P*-value method of hypothesis testing, as summarized in Figure 7-9.

Step 1: The claim of a lower flu rate for vaccinated children can be represented by $p_1 < p_2$.

Step 2: If $p_1 < p_2$ is false, then $p_1 \geq p_2$.

Step 3: Because our claim of $p_1 < p_2$ does not contain equality, it becomes the alternative hypothesis. The null hypothesis is the statement of equality, so we have

$$H_0: p_1 = p_2 \qquad H_1: p_1 < p_2 \text{ (original claim)}$$

Step 4: The significance level is $\alpha = 0.05$.

Step 5: We will use the normal distribution (with the test statistic previously given) as an approximation to the binomial distribution.

Step 6: The test statistic uses the values of $\bar{p}$ and $\bar{q}$, which are as follows:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{14 + 95}{1070 + 532} = 0.068040$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.068040 = 0.931960$$

We can now find the value of the test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\bar{p}\,\bar{q}}{n_1} + \dfrac{\bar{p}\,\bar{q}}{n_2}}}$$

$$= \frac{\left(\dfrac{14}{1070} - \dfrac{95}{532}\right) - 0}{\sqrt{\dfrac{(0.068040)(0.931960)}{1070} + \dfrac{(0.068040)(0.931960)}{532}}} = -12.39$$

continued

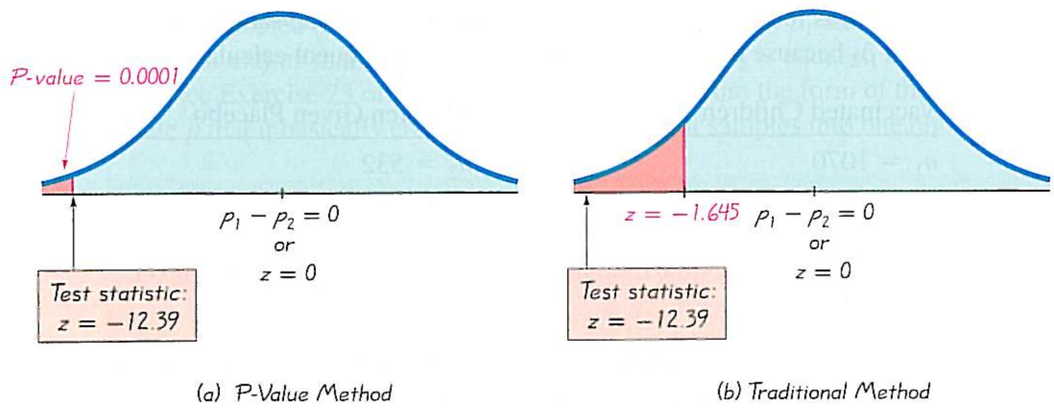

FIGURE 8-1 **Testing Claim that $p_1 < p_2$.**

The P-value of 0.0001 is found as follows: With an alternative hypothesis H_1: $p_1 < p_2$, this is a left-tailed test, so the P-value is the area to the left of the test statistic $z = -12.39$. (See Figure 7-6 which summarizes the procedures for finding P-values in left-tailed tests, right-tailed tests, and two-tailed tests.) Refer to Table A-2 and find that the area to the left of the test statistic $z = -12.39$ is 0.0001, so the P-value is 0.0001. (In Table A-2 we use 0.0001 for the area to the left of any z score of -3.50 or less, but software shows that a more exact P-value is considerably smaller than 0.0001.) The test statistic and P-value are shown in Figure 8-1(a).

Step 7: Because the P-value of 0.0001 is less than the significance level of $\alpha = 0.05$, we reject the null hypothesis of $p_1 = p_2$.

INTERPRETATION We must address the original claim that children given the vaccine developed flu at a rate that is less than the rate for those in the placebo group. Because we reject the null hypothesis, we support the alternative hypothesis and conclude that there is sufficient evidence to support the claim of a lower flu rate for the vaccinated children. (See Figure 7-7 for help in wording the final conclusion.)

Traditional Method of Testing Hypotheses

The preceding example illustrates the P-value approach to hypothesis testing, but it would be quite easy to use the traditional approach (or "critical value approach") instead. In Step 6, instead of finding the P-value, we would find the critical value. With a significance level of $\alpha = 0.05$ in a left-tailed test based on the normal distribution, refer to Table A-2 to find that an area of $\alpha = 0.05$ in the left tail corresponds to the critical value of $z = -1.645$. See Figure 8-1(b) where we can see that the test statistic does fall in the critical region bounded by the critical value of $z = -1.645$. We again reject the null hypothesis. Again, we conclude that there is sufficient evidence to support the claim that vaccinated children developed flu at a rate that is less than the flu rate for children from the placebo group.

Confidence Intervals

We can construct a confidence interval estimate of the difference between population proportions $(p_1 - p_2)$ by using the format given below. If a confidence interval estimate of $p_1 - p_2$ does not include zero, we have evidence suggesting that p_1 and p_2 have different values. (If p_1 and p_2 are the same value, then $p_1 = p_2$, which is equivalent to $p_1 - p_2 = 0$.)

Confidence Interval Estimate of $p_1 - p_2$

The confidence interval estimate of the difference $p_1 - p_2$ is:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

where $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$

EXAMPLE Estimating Vaccine's Effectiveness Use the sample data given in Table 8-1 to construct a 90% confidence interval for the difference between the two population proportions. (The confidence level of 90% is comparable to the significance level of $\alpha = 0.05$ used in the preceding left-tailed hypothesis test. Table 7-2 in Section 7-2 shows that a one-tailed hypothesis test conducted at a 0.05 significance level corresponds to a 90% confidence interval.)

SOLUTION

REQUIREMENT ✓ We should first verify that the necessary requirements (listed earlier in the section) are satisfied. See the preceding example for the verification that the requirements are satisfied. The same verification applies to this example, and we can now proceed to construct the confidence interval. ✓

With a 90% confidence level, $z_{\alpha/2} = 1.645$ (from Table A-2). We first calculate the value of the margin of error E as shown:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.645\sqrt{\frac{\left(\frac{14}{1070}\right)\left(\frac{1056}{1070}\right)}{1070} + \frac{\left(\frac{95}{532}\right)\left(\frac{437}{532}\right)}{532}}$$

$$= 0.027906$$

With $\hat{p}_1 = 14/1070 = 0.013084$, $\hat{p}_2 = 95/532 = 0.178571$, and $E = 0.027906$, the confidence interval is evaluated as follows:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

$$(0.013084 - 0.178571) - 0.027906 < (p_1 - p_2) < (0.013084 - 0.178571) + 0.027906$$

$$-0.193 < (p_1 - p_2) < -0.138$$

continued

INTERPRETATION Although the resulting confidence interval is technically correct, its use of negative values for both confidence interval limits causes it to be somewhat unclear. It would be better to say that the vaccine treatment group developed flu at a lower rate than the placebo group, and the difference appears to be between 0.138 and 0.193. Also, the confidence interval limits do not contain zero, suggesting that there *is* a significant difference between the two proportions. (Because the value of 0 is not a likely value of the difference, it is unlikely that the two proportions are equal.)

Rationale: Why Do the Procedures of This Section Work? The test statistic given for hypothesis tests is justified by the following:

1. With $n_1 p_1 \geq 5$ and $n_1 q_1 \geq 5$, the distribution of $\hat{p}_1$ can be approximated by a normal distribution with mean p_1 and standard deviation $\sqrt{p_1 q_1 / n_1}$ and variance $p_1 q_1 / n_1$. These conclusions are based on Sections 5-6 and 6-5, and they also apply to the second sample.
2. Because $\hat{p}_1$ and $\hat{p}_2$ are each approximated by a normal distribution, $\hat{p}_1 - \hat{p}_2$ will also be approximated by a normal distribution with mean $p_1 - p_2$ and variance

$$\sigma^2_{(\hat{p}_1 - \hat{p}_2)} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

(The above result is based on this property: The variance of the *differences* between two independent random variables is the *sum* of their individual variances.)

3. Because the values of p_1, q_1, p_2, and q_2 are typically unknown and from the null hypothesis we assume that $p_1 = p_2$, we can pool (or combine) the sample data. The pooled estimate of the common value of p_1 and p_2 is $\bar{p} = (x_1 + x_2)/(n_1 + n_2)$. If we replace p_1 and p_2 by $\bar{p}$ and replace q_1 and q_2 by $\bar{q} = 1 - \bar{p}$, the variance from Step 2 leads to the following standard deviation.

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}$$

4. We now know that the distribution of $p_1 - p_2$ is approximately normal, with mean $p_1 - p_2$ and standard deviation as shown in Step 3, so that the z test statistic has the form given earlier.

The form of the confidence interval requires an expression for the variance different from the one given in Step 3. In Step 3 we are assuming that $p_1 = p_2$, but if we don't make that assumption (as in the construction of a confidence interval), we estimate the variance of $p_1 - p_2$ as

$$\sigma^2_{(\hat{p}_1 - \hat{p}_2)} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}$$

and the standard deviation is estimated by

$$\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

In the test statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

let z be positive and negative (for two tails) and solve for $p_1 - p_2$. The results are the limits of the confidence interval given earlier.

8-2 Exercises

Finding Number of Successes. *In Exercises 1–4, find the number of successes x suggested by the given statement.*

1. From the Arizona Department of Weights and Measures: Among 37 inspections at NAPA Auto Parts stores, 81% failed.

2. From the *New York Times*: Among 240 vinyl gloves subjected to stress tests, 63% leaked.

3. From *Sociological Methods and Research*: When 294 central-city residents were surveyed, 28.9% refused to respond.

4. From a Time/CNN survey: 24% of 205 single women said that they "definitely want to get married."

Calculations for Testing Claims. *In Exercises 5 and 6, assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Use the given sample sizes and numbers of successes to find (a) the pooled estimate $\bar{p}$, (b) the z test statistic, (c) the critical z values, and (d) the P-value.*

5.

Treatment	Placebo
$n_1 = 436$	$n_2 = 121$
$x_1 = 192$	$x_2 = 40$

6.

Low Activity	High Activity
$n_1 = 10{,}239$	$n_2 = 9877$
$x_1 = 101$	$x_2 = 56$

7. Exercise and Heart Disease In a study of women and heart disease, the following sample results were obtained: Among 10,239 women with a low level of physical activity (less than 200 kcal/wk), there were 101 cases of heart disease. Among 9877 women with physical activity measured between 200 and 600 kcal/wk, there were 56 cases of heart disease (based on data from "Physical Activity and Coronary Heart Disease in Women" by Lee, Rexrode, et al., *Journal of the American Medical Association,* Vol. 285, No. 11). Construct a 90% confidence interval estimate for the difference between the two proportions. Does the difference appear to be substantial? Does it appear that physical activity corresponds to a lower rate of heart disease?

8. Exercise and Heart Disease Refer to the sample data in Exercise 7 and use a 0.05 significance level to test the claim that the rate of heart disease is higher for women with the lower levels of physical activity. What does the conclusion suggest?

9. Effectiveness of Smoking Bans The Joint Commission on Accreditation of Healthcare Organizations mandated that hospitals ban smoking by 1994. In a study of the effects of this ban, subjects who smoke were randomly selected from two different populations. Among 843 smoking employees of hospitals with the smoking ban, 56 quit smoking one year after the ban. Among 703 smoking employees from workplaces