9.5 Population Proportion

In this section we show how to conduct a hypothesis test about a popul…
Using p_0 to denote the hypothesized value for the population proportion, …
for a hypothesis test about a population proportion are as follows.

$$H_0: p \geq p_0 \qquad H_0: p \leq p_0 \qquad H_0: p = p_0$$
$$H_a: p < p_0 \qquad H_a: p > p_0 \qquad H_a: p \neq p_0$$

The first form is called a lower tail test, the second form is called an upper … third form is called a two-tailed test.

Hypothesis tests about a population proportion are based on the difference … sample proportion $\bar{p}$ and the hypothesized population proportion p_0. The methods … conduct the hypothesis test are similar to those used for hypothesis tests … mean. The only difference is that we use the sample proportion and its … compute the test statistic. The p-value approach or the critical value approach … to determine whether the null hypothesis should be rejected.

Let us consider an example involving a situation faced by Pine Creek golf … the past year, 20% of the players at Pine Creek were women. In an effort … proportion of women players, Pine Creek implemented a special promotion … attract women golfers. One month after the promotion was implemented, … ager requested a statistical study to determine whether the proportion of … Pine Creek had increased. Because the objective of the study is to determine whether … proportion of women golfers increased, an upper tail test with $H_a: p > .20$ … The null and alternative hypotheses for the Pine Creek hypothesis test are …

$$H_0: p \leq .20$$
$$H_a: p > .20$$

If H_0 can be rejected, the test results will give statistical support for the … proportion of women golfers increased and the promotion was beneficial. The … specified that a level of significance of $\alpha = .05$ be used in carrying out this …

The next step of the hypothesis testing procedure is to select a sample … the value of an appropriate test statistic. To show how this step is done for the … tail test, we begin with a general discussion of how to compute the value of the … any form of a hypothesis test about a population proportion. The sampling … point estimator of the population parameter p, is the basis for developing the …

When the null hypothesis is true as an equality, the expected value … hypothesized value p_0; that is, $E(\bar{p}) = p_0$. The standard error of $\bar{p}$ is given by

$$\sigma_{\bar{p}} = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

In Chapter 7 we said that if $np \geq 5$ and $n(1 - p) \geq 5$, the sampling … be approximated by a normal distribution.[4] Under these conditions, which … practice, the quantity

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}}$$

[4]In most applications involving hypothesis tests of a population proportion, sample sizes are … mal approximation. The exact sampling distribution of $\bar{p}$ is discrete, with the probability for each … distribution. So hypothesis testing is a bit more complicated for small samples when the normal …

has a standard normal probability distribution. With $\sigma_{\bar{p}} = \sqrt{p_0(1 - p_0)/n}$, the standard normal random variable z is the test statistic used to conduct hypothesis tests about a population proportion.

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \tag{9.4}$$

DATA file WomenGolf

We can now compute the test statistic for the Pine Creek hypothesis test. Suppose a random sample of 400 players was selected, and that 100 of the players were women. The proportion of women golfers in the sample is

$$\bar{p} = \frac{100}{400} = .25$$

Using equation (9.4), the value of the test statistic is

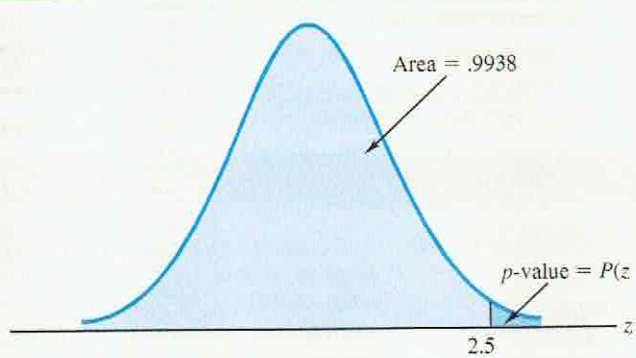
$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{.25 - .20}{\sqrt{\frac{.20(1 - .20)}{400}}} = \frac{.05}{.02} = 2.50$$

Because the Pine Creek hypothesis test is an upper tail test, the p-value is the probability that z is greater than or equal to $z = 2.50$; that is, it is the upper tail area corresponding to $z \geq 2.50$. Using the standard normal probability table, we find that the lower tail area for $z = 2.50$ is .9938. Thus, the p-value for the Pine Creek test is $1.0000 - .9938 = .0062$. Figure 9.9 shows this p-value calculation.

Recall that the course manager specified a level of significance of $\alpha = .05$. A p-value $= .0062 < .05$ gives sufficient statistical evidence to reject H_0 at the .05 level of significance. Thus, the test provides statistical support for the conclusion that the special promotion increased the proportion of women players at the Pine Creek golf course.

The decision whether to reject the null hypothesis can also be made using the critical value approach. The critical value corresponding to an area of .05 in the upper tail of a normal probability distribution is $z_{.05} = 1.645$. Thus, the rejection rule using the critical value approach is to reject H_0 if $z \geq 1.645$. Because $z = 2.50 > 1.645$, H_0 is rejected.

FIGURE 9.9 CALCULATION OF THE p-VALUE FOR THE PINE CREEK HYPOTHESIS TEST



Again, we see that the p-value approach and the critical value approach lead to the same hypothesis testing conclusion, but the p-value approach provides more information. With a p-value $= .0062$, the null hypothesis would be rejected for any level of significance greater than or equal to .0062.

Using Excel

Excel can be used to conduct one-tailed and two-tailed hypothesis tests about a population proportion using the p-value approach. The procedure is similar to the approach used with Excel in conducting hypothesis tests about a population mean. The primary difference is that the test statistic is based on the sampling distribution of $\bar{x}$ for hypothesis tests about a population mean and on the sampling distribution of $\bar{p}$ for hypothesis tests about a population proportion. Thus, although different formulas are used to compute the test statistic needed to make the hypothesis testing decision, the computations of the critical value and the p-value for the tests are identical.

We will illustrate the procedure by showing how Excel can be used to conduct the upper tail hypothesis test for the Pine Creek golf course study. Refer to Figure 9.10 as we describe the tasks involved. The formula worksheet is in the background; the value worksheet is in the foreground.

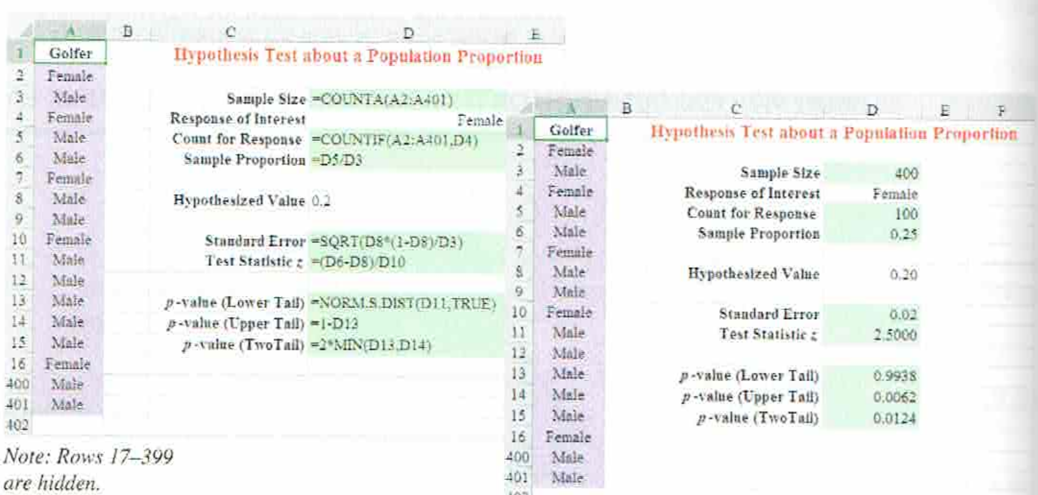
Enter/Access Data: Open the DATAfile named *WomenGolf*. A label and the gender of each golfer in the study are entered into cells A1:A401.

Enter Functions and Formulas: The sample size, response count, and sample proportion are calculated in cells D3, D5, and D6. Because the data are not numeric, Excel's COUNTA function, not the COUNT function, is used in cell D3 to determine the sample size. We entered Female in cell D4 to identify the response for which we wish to compute a proportion. The COUNTIF function is then used in cell D5 to determine the number of responses of the type identified in cell D4. The sample proportion is then computed in cell D6 by dividing the response count by the sample size.

The hypothesized value of the population proportion (.20) is entered into cell D8. The standard error is obtained in cell D10 by entering the formula =SQRT(D8*(1-D8)/D3). The formula =(D6-D8)/D10 entered into cell D11 computes the test statistic z (2.50). To compute the p-value for a lower tail test, we enter the formula =NORM.S.DIST(D11,TRUE)

FIGURE 9.10 EXCEL WORKSHEET: HYPOTHESIS TEST FOR PINE CREEK GOLF COURSE

DATA file
WomenGolf



Note: Rows 17–399 are hidden.

TABLE 9.4 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: p \geq p_0$ $H_a: p < p_0$	$H_0: p \leq p_0$ $H_a: p > p_0$	$H_0: p = p_0$ $H_a: p \neq p_0$
Test Statistic	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

into cell D13. The p-value for an upper tail test is then computed in cell D14 as 1 minus the p-value for the lower tail test. Finally, the p-value for a two-tailed test is computed in cell D15 as two times the minimum of the two one-tailed p-values. The value worksheet shows that the three p-values are as follows: p-value (Lower Tail) = 0.9938; p-value (Upper Tail) = 0.0062; and p-value (Two Tail) = 0.0124.

The development of the worksheet is now complete. For the Pine Creek upper tail hypothesis test, we reject the null hypothesis that the population proportion is .20 or less because the p-value (Upper Tail) = 0.0062 is less than $\alpha = .05$. Indeed, with this p-value we would reject the null hypothesis for any level of significance of .0062 or greater.

A template for other problems The worksheet in Figure 9.10 can be used as a template for hypothesis tests about a population proportion whenever $np \geq 5$ and $n(1 - p) \geq 5$. Just enter the appropriate data in column A, adjust the ranges for the formulas in cells D3 and D5, enter the appropriate response in cell D4, and enter the hypothesized value in cell D8. The standard error, the test statistic, and the three p-values will then appear. Depending on the form of the hypothesis test (lower tail, upper tail, or two-tailed), we can then choose the appropriate p-value to make the rejection decision.

Summary

The procedure used to conduct a hypothesis test about a population proportion is similar to the procedure used to conduct a hypothesis test about a population mean. Although we only illustrated how to conduct a hypothesis test about a population proportion for an upper tail test, similar procedures can be used for lower tail and two-tailed tests. Table 9.4 provides a summary of the hypothesis tests about a population proportion. We assume that $np \geq 5$ and $n(1 - p) \geq 5$; thus the normal probability distribution can be used to approximate the sampling distribution of $\bar{p}$.

NOTE AND COMMENT

When testing a hypothesis of the population proportion with a sample size that is at least 5% of the population size (that is, $n/N \geq .05$), the finite population correction factor should be used when calculating the standard error of the sampling distribution of $\bar{p}$, i.e., $s_{\bar{p}} = \sqrt{\dfrac{N - n}{N - 1}}\sqrt{\dfrac{p_0(1 - p_0)}{n}}$.

Exercises

Methods

35. Consider the following hypothesis test:

$$H_0: p = .20$$
$$H_a: p \neq .20$$

A sample of 400 provided a sample proportion $\bar{p} = .175$.

a. Compute the value of the test statistic.
b. What is the p-value?
c. At $\alpha = .05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

SELF*test*

36. Consider the following hypothesis test:

$$H_0: p \geq .75$$
$$H_a: p < .75$$

A sample of 300 items was selected. Compute the p-value and state your conclusion for each of the following sample results. Use $\alpha = .05$.

a. $\bar{p} = .68$
b. $\bar{p} = .72$
c. $\bar{p} = .70$
d. $\bar{p} = .77$

Applications

37. The U.S. Bureau of Labor Statistics reports that 11.3% of U.S. workers belong to unions (BLS website, January 2014). Suppose a sample of 400 U.S. workers is collected in 2014 to determine whether union efforts to organize have increased union membership.
    a. Formulate the hypotheses that can be used to determine whether union membership increased in 2014.
    b. If the sample results show that 52 of the workers belonged to unions, what is the p-value for your hypothesis test?
    c. At $\alpha = .05$, what is your conclusion?

SELF*test*

38. A study by *Consumer Reports* showed that 64% of supermarket shoppers believe supermarket brands to be as good as national name brands. To investigate whether this result applies to its own product, the manufacturer of a national name-brand ketchup asked a sample of shoppers whether they believed that supermarket ketchup was as good as the national brand ketchup.
    a. Formulate the hypotheses that could be used to determine whether the percentage of supermarket shoppers who believe that the supermarket ketchup was as good as the national brand ketchup differed from 64%.
    b. If a sample of 100 shoppers showed 52 stating that the supermarket brand was as good as the national brand, what is the p-value?
    c. At $\alpha = .05$, what is your conclusion?
    d. Should the national brand ketchup manufacturer be pleased with this conclusion? Explain.

DATA*file*
HomeState

39. What percentage of the population live in their state of birth? According to the U.S. Census Bureau's American Community Survey, the figure ranges from 25% in Nevada to 78.7% in Louisiana (*AARP Bulletin*, March 2014). The average percentage across all states and the District of Columbia is 57.7%. The data in the DATAfile *Homestate* are consistent with the findings in the American Community Survey. The data represent a random sample of 120 Arkansas residents and for a random sample of 180 Virginia residents.