8.5: Large Sample Tests for a Population Proportion

Learning Objectives

- To learn how to apply the five-step critical value test procedure for test of hypotheses concerning a population proportion.
- To learn how to apply the five-step p-value test procedure for test of hypotheses concerning a population proportion.

Both the critical value approach and the p-value approach can be applied to test hypotheses about a population proportion p. The null hypothesis will have the form $H_0 : p = p_0$ for some specific number p_0 between 0 and 1. The alternative hypothesis will be one of the three inequalities

- $p <p_0$,
- $p>p_0$, or
- $p≠p_0$

for the same number p_0 that appears in the null hypothesis.

The information in Section 6.3 gives the following formula for the test statistic and its distribution. In the formula p_0 is the numerical value of p that appears in the two hypotheses, $q_0=1−p_0, \hat{p}$ is the sample proportion, and n is the sample size. Remember that the condition that the sample be large is not that n be at least 30 but that the interval

$$\left[\hat{p} −3 \sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} , \hat{p} + 3 \sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} \right] \nonumber$$

lie wholly within the interval $[0,1]$.

Standardized Test Statistic for Large Sample Hypothesis Tests Concerning a Single Population Proportion

$$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0q_o}{n}}} \tag{eq2}$$

The test statistic has the standard normal distribution.

The distribution of the standardized test statistic and the corresponding rejection region for each form of the alternative hypothesis (left-tailed, right-tailed, or two-tailed), is shown in Figure \(\PageIndex{1}\).

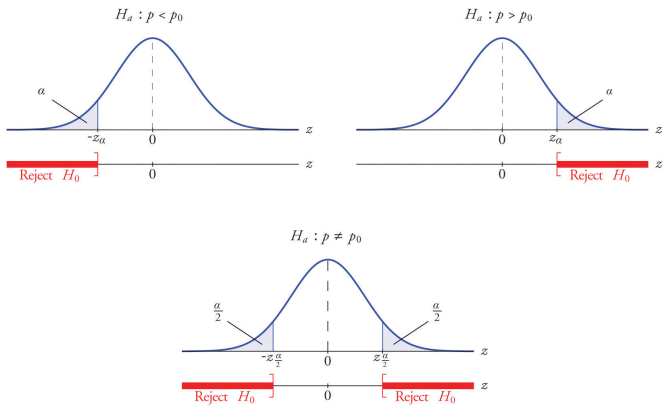

Figure \(\PageIndex{1}\): Distribution of the Standardized Test Statistic and the Rejection Region

Example \(\PageIndex{1}\)

A soft drink maker claims that a majority of adults prefer its leading beverage over that of its main competitor's. To test this claim \(500\) randomly selected people were given the two beverages in random order to taste. Among them, \(270\) preferred the soft drink maker's brand, \(211\) preferred the competitor's brand, and \(19\) could not make up their minds. Determine whether there is sufficient evidence, at the \(5\%\) level of significance, to support the soft drink maker's claim against the default that the population is evenly split in its preference.

Solution

We will use the critical value approach to perform the test. The same test will be performed using the \(p\)-value approach in Example \(\PageIndex{3}\).

We must check that the sample is sufficiently large to validly perform the test. Since \(\hat{p} =270/500=0.54\),
$$\sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} =\sqrt{ \dfrac{(0.54)(0.46)}{500}} ≈0.02 \nonumber$$

hence

$$\begin{align} & \left[\hat{p} −3\sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} ,\hat{p} +3\sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} \right] \\ &=[0.54−(3)(0.02),0.54+(3)(0.02)] \\ &=[0.48, 0.60] ⊂[0,1] \end{align} \nonumber$$

so the sample is sufficiently large.

- **Step 1.** The relevant test is

$$H_0 : p = 0.50 \quad \nonumber$$

$$vs. \nonumber$$

$$H_a : p > 0.50, @ \,\alpha =0.05 \nonumber$$

where p denotes the proportion of all adults who prefer the company's beverage over that of its competitor's beverage.

- **Step 2.** The test statistic (Equation \ref{eq2}) is

$$Z=\dfrac{\hat{p} -p_0}{\sqrt{ \dfrac{p_0q_0}{n}}} \nonumber$$

and has the standard normal distribution.

- **Step 3.** The value of the test statistic is

$$\begin{align} Z &=\dfrac{\hat{p} -p_0}{\sqrt{ \dfrac{p_0q_0}{n}}} \\[6pt] &= \dfrac{0.54-0.50}{\sqrt{\dfrac{(0.50)(0.50)}{500}}} \\[6pt] &=1.789 \end{align} \nonumber$$

- **Step 4.** Since the symbol in H_a is "$>$" this is a right-tailed test, so there is a single critical value, $z_{α}=z_{0.05}$. Reading from the last line in Figure 7.1.6 its value is 1.645. The rejection region is $[1.645,∞)$.
- **Step 5.** As shown in Figure $\PageIndex{2}$ the test statistic falls in the rejection region. The decision is to reject H_0. In the context of the problem our conclusion is:

  The data provide sufficient evidence, at the 5% level of significance, to conclude that a majority of adults prefer the company's beverage to that of their competitor's.

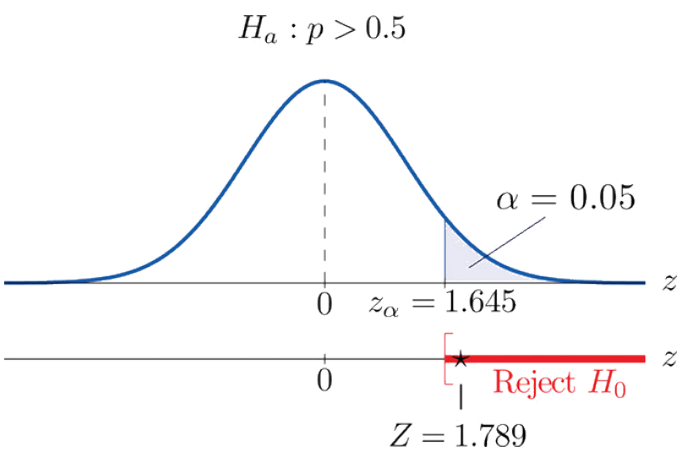

Figure $\PageIndex{2}$: Rejection Region and Test Statistic for Example $\PageIndex{1}$

Example $\PageIndex{2}$

Globally the long-term proportion of newborns who are male is 51.46%. A researcher believes that the proportion of boys at birth changes under severe economic conditions. To test this belief randomly selected birth records of $5,000$ babies born during a period of economic recession were examined. It was found in the sample that 52.55% of the

newborns were boys. Determine whether there is sufficient evidence, at the 10% level of significance, to support the researcher's belief.

Solution

We will use the critical value approach to perform the test. The same test will be performed using the p-value approach in Example $\PageIndex{1}$.

The sample is sufficiently large to validly perform the test since

$$\sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} =\sqrt{ \dfrac{(0.5255)(0.4745)}{5000}} ≈0.01 \nonumber$$

hence

$$\begin{align} & \left[\hat{p} −3\sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} ,\hat{p} +3\sqrt{ \dfrac{\hat{p} (1−\hat{p})}{n}} \right] \\ &=[0.5255−0.03,0.5255+0.03] \\ &=[0.4955,0.5555] ⊂[0,1] \end{align} \nonumber$$

- **Step 1**. Let p be the true proportion of boys among all newborns during the recession period. The burden of proof is to show that severe economic conditions change it from the historic long-term value of 0.5146 rather than to show that it stays the same, so the hypothesis test is

$$H_0 : p = 0.5146 \nonumber$$

$$vs. \nonumber$$

$$H_a : p \neq 0.5146\, @ \,\alpha =0.10 \nonumber$$

- **Step 2.** The test statistic (Equation \ref{eq2}) is

$$Z=\dfrac{\hat{p} −p_0}{\sqrt{ \dfrac{p_0q_0}{n}}} \nonumber$$

and has the standard normal distribution.

- **Step 3.** The value of the test statistic is

$$\begin{align} Z &=\dfrac{\hat{p} −p_0}{\sqrt{ \dfrac{p_0q_0}{n}}} \\[6pt] &= \dfrac{0.5255−0.5146}{\sqrt{\dfrac{(0.5146)(0.4854)}{5000}}} \\[6pt] &=1.542 \end{align} \nonumber$$

- **Step 4.** Since the symbol in H_a is "$\neq$" this is a two-tailed test, so there are a pair of critical values, $\pm z_{\alpha /2}=\pm z_{0.05}=\pm 1.645$. The rejection region is $(-\infty ,-1.645]\cup [1.645,\infty)$.
- **Step 5.** As shown in Figure $\PageIndex{3}$ the test statistic does not fall in the rejection region. The decision is not to reject H_0. In the context of the problem our conclusion is:

  The data do not provide sufficient evidence, at the 10% level of significance, to conclude that the proportion of newborns who are male differs from the historic proportion in times of economic recession.

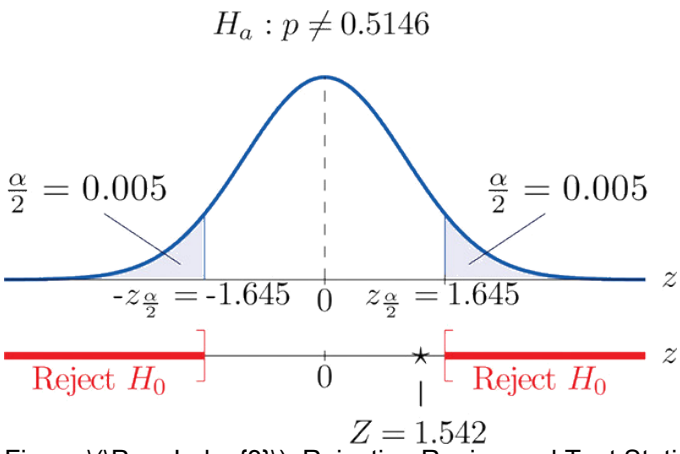

Figure \(\PageIndex{3}\): Rejection Region and Test Statistic for Example \(\PageIndex{2}\).

Example \(\PageIndex{3}\)

Perform the test of Example \(\PageIndex{1}\) using the \(p\)-value approach.

Solution

We already know that the sample size is sufficiently large to validly perform the test.

- **Steps 1–3** of the five-step procedure described in Section 8.3 have already been done in Example \(\PageIndex{1}\) so we will not repeat them here, but only say that we know that the test is right-tailed and that value of the test statistic is \(Z = 1.789\).
- **Step 4.** Since the test is right-tailed the p-value is the area under the standard normal curve cut off by the observed test statistic, \(Z = 1.789\), as illustrated in Figure \(\PageIndex{4}\). By Figure 7.1.5 that area and therefore the p-value is \(1−0.9633=0.0367\).
- **Step 5.** Since the \(p\)-value is less than \(α=0.05\) the decision is to reject \(H_0\).

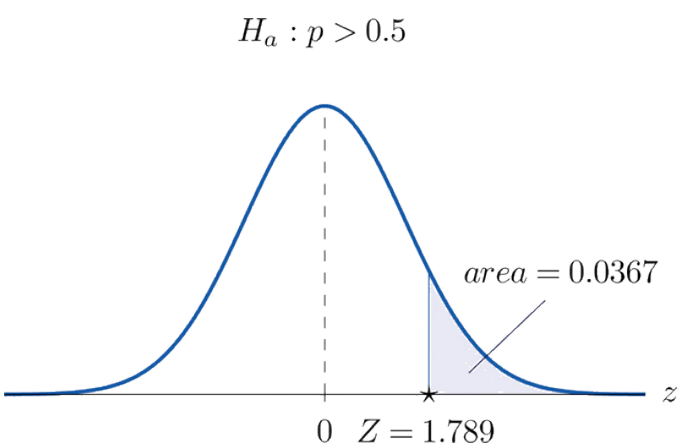

Figure \(\PageIndex{4}\): P-Value for Example \(\PageIndex{3}\)

Example \(\PageIndex{4}\)

Perform the test of Example \(\PageIndex{2}\) using the \(p\)-value approach.

Solution

We already know that the sample size is sufficiently large to validly perform the test.

- **Steps 1–3** of the five-step procedure described in Section 8.3 have already been done in Example \(\PageIndex{2}\). They tell us that the test is two-tailed and that value of the test statistic is \(Z = 1.542\).
- **Step 4.** Since the test is two-tailed the \(p\)-value is the double of the area under the standard normal curve cut off by the observed test statistic, \(Z = 1.542\). By Figure 7.1.5 that area is \(1-0.9382=0.0618\), as illustrated in Figure \(\PageIndex{5}\), hence the \(p\)-value is \(2\times 0.0618=0.1236\).
- **Step 5.** Since the \(p\)-value is greater than \(\alpha =0.10\) the decision is not to reject \(H_0\).

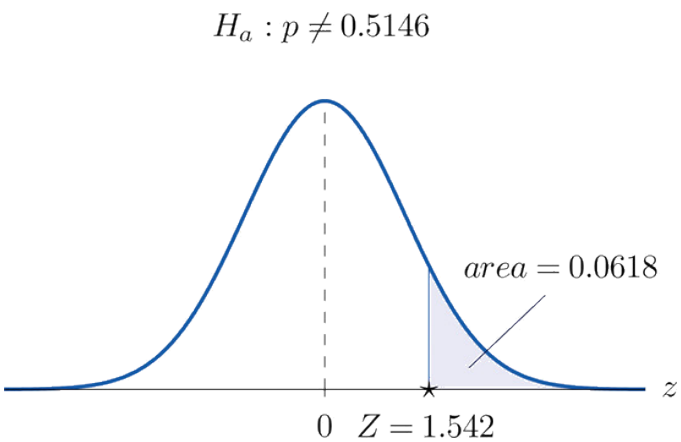

Figure \(\PageIndex{5}\): P-Value for Example \(\PageIndex{4}\)

Key Takeaway

- There is one formula for the test statistic in testing hypotheses about a population proportion. The test statistic follows the standard normal distribution.
- Either five-step procedure, critical value or \(p\)-value approach, can be used.