Testing the Difference Between Means Proportions

OBJECTIVES

How to perform a z-test for the difference between two population proportions p_1 *and* p_2

null and alternative hypotheses

$$\begin{cases} H_0: p_1 = p_2 \\ H_a: p_1 \neq p_2 \end{cases}, \quad \begin{cases} H_0: p_1 \leq p_2 \\ H_a: p_1 > p_2 \end{cases}, \quad \text{and} \quad \begin{cases} H_0: p_1 \geq p_2 \\ H_a: p_1 < p_2 \end{cases}.$$

Regardless of which hypotheses you use, you always assume there is no difference between the population proportions, or $p_1 = p_2$.

For instance, suppose you want to determine whether the proportion of female college students who earn a bachelor's degree in four years is different from the proportion of male college students who earn a bachelor's degree in four years. The following conditions are necessary to use a z-test to test such a difference.

1. The samples must be randomly selected.
2. The samples must be independent.
3. The samples must be large enough to use a normal sampling distribution. That is, $n_1 p_1 \geq 5$, $n_1 q_1 \geq 5$, $n_2 p_2 \geq 5$, and $n_2 q_2 \geq 5$.

The standardized test statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The standard error

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

The weighted estimate

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

THE TWO-SAMPLE *t*-TEST FOR THE DIFFERENCE BETWEEN PROPORTIONS

If the sampling distribution for $\hat{p}_1 - \hat{p}_2$ is normal, you can use a two-sample z-test to test the difference between two population proportions p_1 and p_2.

TWO-SAMPLE *z*-TEST FOR THE DIFFERENCE BETWEEN PROPORTIONS

A two-sample z-test is used to test the difference between two population proportions p_1 and p_2 when a sample is randomly selected from each population. The **test statistic** is $\hat{p}_1 - \hat{p}_2$, and the **standardized test statistic** is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{and} \quad \overline{q} = 1 - \overline{p}.$$

Note: $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ must be at least 5.

If the null hypothesis states $p_1 = p_2$, $p_1 \leq p_2$, or $p_1 \geq p_2$, then $p_1 = p_2$ is assumed and the expression $p_1 - p_2$ is equal to 0 in the preceding test.

STUDY TIP

The following symbols are used in the z-test for $p_1 - p_2$. See Sections 4.2 and 5.5 to review the binomial distribution.

Symbol	Description
p_1, p_2	Population proportions
x_1, x_2	Number of successes in each sample
n_1, n_2	Size of each sample
$\hat{p}_1, \hat{p}_2$	Sample proportions of successes
$\overline{p}$	Weighted estimate for p_1 and p_2

THE TWO-SAMPLE *t*-TEST FOR THE DIFFERENCE BETWEEN PROPORTIONS

A hypothesis test for the difference between proportions can also be performed using *P*-values. Use the guidelines listed above, skipping Steps 3 and 4. After finding the standardized test statistic, use the Standard Normal Table to calculate the *P*-value. Then make a decision to reject or fail to reject the null hypothesis. If P is less than or equal to α, reject H_0. Otherwise, fail to reject H_0.

GUIDELINES

Using a Two-Sample *z*-Test for the Difference Between Proportions

IN WORDS	IN SYMBOLS
1. State the claim mathematically and verbally. Identify the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Determine the critical value(s).	Use Table 4 in Appendix B.
4. Determine the rejection region(s).	
5. Find the weighted estimate of p_1 and p_2. Verify that $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ are at least 5.	$\overline{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$
6. Find the standardized test statistic and sketch the sampling distribution.	$z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$
7. Make a decision to reject or fail to reject the null hypothesis.	If z is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

Testing the Difference Between Means Proportions

EXAMPLE 1 A Two-Sample z-Test for the Difference Between Proportions

A study of 150 randomly selected occupants in passenger cars and 200 randomly selected occupants in pickup trucks shows that 86% of occupants in passenger cars and 74% of occupants in pickup trucks wear seat belts. At $\alpha = 0.10$, can you reject the claim that the proportion of occupants who wear seat belts is the same for passenger cars and pickup trucks? *(Adapted from National Highway Traffic Safety Administration)*

Sample Statistics for Vehicles

Passenger cars	Pickup trucks
$n_1 = 150$	$n_2 = 200$
$\hat{p}_1 = 0.86$	$\hat{p}_2 = 0.74$
$x_1 = 129$	$x_2 = 148$

Solution

The claim is "the proportion of occupants who wear seat belts is the same for passenger cars and pickup trucks." So, the null and alternative hypotheses are

$H_0: p_1 = p_2$ (Claim) and $H_a: p_1 \neq p_2$.

Because the test is two-tailed and the level of significance is $\alpha = 0.10$, the critical values are $-z_0 = -1.645$ and $z_0 = 1.645$. The rejection regions are $z < -1.645$ and $z > 1.645$. The weighted estimate of p_1 and p_2 is

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{129 + 148}{150 + 200} = \frac{277}{350} \approx 0.7914$$

and

$$\overline{q} = 1 - \overline{p} \approx 1 - 0.7914 = 0.2086.$$

Because $n_1\overline{p} \approx 150(0.7914)$, $n_1\overline{q} \approx 150(0.2086)$, $n_2\overline{p} \approx 200(0.7914)$, and $n_2\overline{q} \approx 200(0.2086)$ are at least 5, you can use a two-sample z-test. The standardized test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \frac{(0.86 - 0.74) - 0}{\sqrt{(0.7914)(0.2086)\left(\frac{1}{150} + \frac{1}{200}\right)}} \approx 2.73.$$

The graph at the left shows the location of the rejection regions and the standardized test statistic. Because z is in the rejection region, you should decide to reject the null hypothesis.

Interpretation There is enough evidence at the 10% level of significance to reject the claim that the proportion of occupants who wear seat belts is the same for passenger cars and pickup trucks.

The graph at the right shows the location of the rejection region and the standardized test statistic t. Because t is in the rejection region, you should decide to reject the null hypothesis.

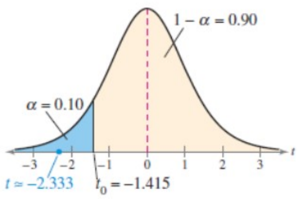
Interpretation There is enough evidence at the 10% level of significance to support the shoe manufacturer's claim that athletes can increase their vertical jump heights using the new Strength Shoes®.

The standardized test statistic is

$$t = \frac{\overline{d} - \mu_d}{s_d/\sqrt{n}} \quad \text{Use the } t\text{-test.}$$

$$\approx \frac{-1.75 - 0}{2.1213/\sqrt{8}} \quad \text{Assume } \mu_d = 0.$$

$$\approx -2.333.$$



Testing the Difference Between Means Proportions

Try It Yourself 1

Consider the results of the *NYTS* study discussed in the Chapter Opener. At $\alpha = 0.05$, can you support the claim that there is a difference between the proportion of male high school students who smoke cigarettes and the proportion of female high school students who smoke cigarettes?

a. Identify the *claim* and state H_0 and H_a.
b. Identify the *level of significance* α.
c. Find the *critical values* and identify the *rejection regions*.
d. *Find* $\overline{p}$ and $\overline{q}$.
e. *Verify* that $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ are at least 5.
f. Find the *standardized test statistic z*. *Sketch* a graph.
g. *Decide* whether to reject the null hypothesis.
h. *Interpret* the decision in the context of the original claim.

Testing the Difference Between Means Proportions

EXAMPLE 2 A Two-Sample z-Test for the Difference Between Proportions

A medical research team conducted a study to test the effect of a cholesterol-reducing medication. At the end of the study, the researchers found that of the 4700 randomly selected subjects who took the medication, 301 died of heart disease. Of the 4300 randomly selected subjects who took a placebo, 357 died of heart disease. At $\alpha = 0.01$, can you support the claim that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo? *(Adapted from The New England Journal of Medicine)*

Sample Statistics for Cholesterol-Reducing Medication

Received medication	Received placebo
$n_1 = 4700$	$n_2 = 4300$
$x_1 = 301$	$x_2 = 357$
$\hat{p}_1 = 0.064$	$\hat{p}_2 = 0.083$

Solution

The claim is "the death rate due to heart disease is lower for those who took the medication than for those who took the placebo." So, the null and alternative hypotheses are

$H_0: p_1 \geq p_2$ and $H_a: p_1 < p_2$. (Claim)

Because the test is left-tailed and the level of significance is $\alpha = 0.01$, the critical value is $z_0 = -2.33$. The rejection region is $z < -2.33$. The weighted estimate of p_1 and p_2 is

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{301 + 357}{4700 + 4300} = \frac{658}{9000} \approx 0.0731$$

and

$$\overline{q} = 1 - \overline{p} \approx 1 - 0.0731 = 0.9269.$$

Because $n_1\overline{p} = 4700(0.0731)$, $n_1\overline{q} = 4700(0.9269)$, $n_2\overline{p} = 4300(0.0731)$, and $n_2\overline{q} = 4300(0.9269)$ are at least 5, you can use a two-sample z-test.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \frac{(0.064 - 0.083) - 0}{\sqrt{(0.0731)(0.9269)\left(\frac{1}{4700} + \frac{1}{4300}\right)}} \approx -3.46$$

The graph at the left shows the location of the rejection region and the standardized test statistic. Because z is in the rejection region, you should decide to reject the null hypothesis.

Interpretation There is enough evidence at the 1% level of significance to support the claim that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo.

Testing the Difference Between Means Proportions

Try It Yourself 2

Consider the results of the *NYTS* study discussed in the Chapter Opener. At $\alpha = 0.05$, can you support the claim that the proportion of male high school students who smoke cigars is greater than the proportion of female high school students who smoke cigars?

a. Identify the *claim* and state H_0 and H_a.
b. Identify the *level of significance* α.
c. Find the *critical value* and identify the *rejection region*.
d. *Find* $\overline{p}$ and $\overline{q}$.
e. *Verify* that $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ are at least 5.
f. Find the *standardized test statistic* z. *Sketch* a graph.
g. *Decide* whether to reject the null hypothesis.
h. *Interpret* the decision in the context of the original claim.

Testing the Difference Between Means Proportions

CLASSWORK

PAGE 475

#1 - 16

OBJECTIVES

How to perform a *z*-test for the difference between two population proportions p_1 *and* p_2

null and alternative hypotheses

$$\begin{cases} H_0: p_1 = p_2 \\ H_a: p_1 \neq p_2 \end{cases}, \quad \begin{cases} H_0: p_1 \leq p_2 \\ H_a: p_1 > p_2 \end{cases}, \quad \text{and} \quad \begin{cases} H_0: p_1 \geq p_2 \\ H_a: p_1 < p_2 \end{cases}.$$

Regardless of which hypotheses you use, you always assume there is no difference between the population proportions, or $p_1 = p_2$.

For instance, suppose you want to determine whether the proportion of female college students who earn a bachelor's degree in four years is different from the proportion of male college students who earn a bachelor's degree in four years. The following conditions are necessary to use a z-test to test such a difference.

1. The samples must be randomly selected.
2. The samples must be independent.
3. The samples must be large enough to use a normal sampling distribution. That is, $n_1 p_1 \geq 5$, $n_1 q_1 \geq 5$, $n_2 p_2 \geq 5$, and $n_2 q_2 \geq 5$.

The standardized test statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The standard error

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

The weighted estimate

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