7.2-SJ: Statistical Tests for the Difference in Two Population Proportions
Race and Wrongful Convictions

INTRODUCTION

In the previous lesson, we began testing hypotheses that compare two population proportions. We learned that when the sample size is large enough, the test statistic

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

- has an approximately normal distribution, and
- may be used to test claims regarding the equality of population proportions.

Our approach was informal. In this lesson we will discuss the topic of hypothesis testing in greater detail.

TRY THESE

The United States judicial system, which evaluates laws, is imperfect. Sometimes it makes mistakes when applying laws to individual cases. Some people are convicted of crimes but later exonerated because they were actually innocent. (An **exoneration** occurs when a person who has been convicted of a crime is officially cleared based on new evidence of innocence.) The National Registry of Exonerations (a project of the University of California Irvine, Newkirk Center for Science and Society, the University of Michigan Law School, and the Michigan State University College of Law) collects information about all known exonerations of innocent criminal defendants in the United States, from 1989 to the present.

1 If laws are applied equally, should mistakes also be made equally? In other words, should the exoneration rate be equal for everyone? How would you expect the exoneration rate to differ between African-American criminal defendants and white criminal defendants?

Suppose that we decide to test the claim that the proportion of exonerations among African-Americans convicted of federal crimes is higher than the proportion of exonerations among whites convicted of federal crimes. We gather independent samples. Our data tells us that out of a random sample of n_1 = 10,000 African-Americans convicted of federal crimes, x_1 = 43 were later exonerated. In addition, out of a random sample of n_2 = 10,000 whites convicted of federal crimes, x_2 = 7 were later exonerated.

Using a 1% significance level, we can now test the claim that the true population proportions are different.

Step 1: Determine the Hypotheses

When testing for the difference between two population proportions, the null hypothesis states that the population proportions are equal:

H_0: $p_1 = p_2$ (or, equivalently, H_0: $p_1 - p_2 = 0$)

The alternative hypothesis is one of the following inequalities:

H_a: $p_1 < p_2$	For a left-tailed test
H_a: $p_1 > p_2$	For a right-tailed test
H_a: $p_1 \neq p_2$	For a two-tailed test

2 For the exoneration rate test, we claimed that the proportion of exonerations was higher among African-Americans convicted of federal crimes than among whites convicted of federal crimes.

Let p_1 represent the proportion of African-American federal convicts who were later exonerated. Then let p_2 represent the proportion of white federal convicts who were later exonerated.

A What is the null hypothesis for this test?

B What is the alternative hypothesis?

C Is this a one-tailed or two-tailed test? Think about the null hypothesis. How do you know that this is a one or two-tailed test?

Step 2: Collect the Data

In order to perform a hypothesis test for the difference in two population proportions, we must make sure that the underlying sampling distributions for each sample are approximately normal.

We will consider the normality criteria satisfied if there are at least 10 successes and 10 failures in each sample.

3 For the exoneration rate test, count the number of successes and failures in each sample. In this test, a success is a person *who has been exonerated of a federal crime.*

A For African-American federal convicts, there are __________ successes and __________ failures.

B For white federal convicts, there are __________ successes and __________ failures.

C Are these successes and failures enough to consider that the approximate normality criteria have been met?

The sample proportions

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

estimate the population proportions, p_1 and p_2, respectively. The difference in the observed sample proportions is $\hat{p}_1 - \hat{p}_2$.

4 For the exoneration rate test:

A Calculate $\hat{p}_1$ and $\hat{p}_2$.

B Calculate the observed difference in sample proportions, $\hat{p}_1 - \hat{p}_2$.

Step 3: Assess the Evidence

We use a sampling distribution of differences between sample proportions to assess the likelihood of the observed difference occurring. As we have seen, the *test statistic* for the two proportions test is

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

This test statistic is a *Z*-score for the difference in the observed sample proportions. To find the *Z*-score, we

- subtract the mean of differences $(p_1 - p_2)$ from the observed difference, and
- divide by the standard error of differences.

The null hypothesis indicates that we should use $p_1 - p_2 = 0$ for the mean of differences. The formula also requires the pooled proportion,

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

5 For the exoneration rate test, the pooled proportion is

$$\bar{p} = \frac{43 + 7}{10{,}000 + 10{,}000} = 0.0025$$

With this, the standard error of differences is

$$s_{\hat{p}_i - \hat{p}_f} = \sqrt{\frac{0.0025(1-0.0025)}{10{,}000} + \frac{0.0025(1-0.0025)}{10{,}000}} \approx 7.06 \text{ E -4} = 0.000706$$

To compute the test statistic for the exoneration rate study, use

- the difference in the observed sample proportions, and
- the estimated standard error.

$$Z = \frac{(\hat{p}_i - \hat{p}_f) - (p_i - p_f)}{s_{\hat{p}_i - \hat{p}_f}} = \frac{(\hat{p}_i - \hat{p}_f)}{s_{\hat{p}_i - \hat{p}_f}}$$

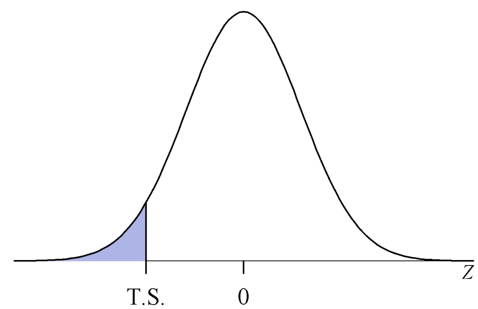
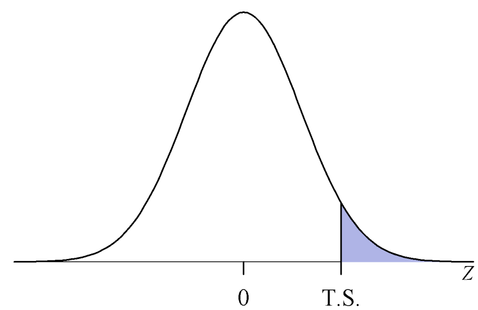
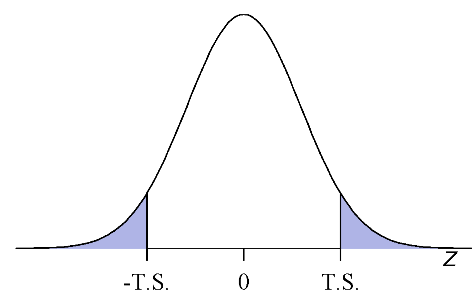
A Calculate the *Z*-score using the observed difference you computed in Question 4B.

B How many standard errors is the observed difference, $\hat{p}_i - \hat{p}_f$, from the *assumed* mean of differences of zero?

C Does this observed difference seem to be unusually far from the assumed mean of differences? Why or why not?

Let's revisit the concept of the *P*-value in a hypothesis test. The *P*-value is the probability of observing a *Z*-score that is at least as extreme as the test statistic, *assuming the null-hypothesis is true*.

Recall that there are three types of hypothesis tests:

For a left-tailed test, where the alternative hypothesis is $H_a: p_1 < p_2$, the *P*-value is the area to the left of the test statistic.	
For a right-tailed test, where the alternative hypothesis is $H_a: p_1 > p_2$, the *P*-value is the area to the right of the test statistic.	
For a two-tailed test, where the alternative hypothesis is $H_a: p_1 \neq p_2$, the *P*-value is • *twice* the area to the right of a positive test statistic, or • *twice* the area to the left of a negative test statistic.	

6 Sketch the standard normal curve. Identify the position of the test statistic. Shade in the area that corresponds to the *P*-value.

7 Write the *P*-value below. Remember that the tail's area is the *P*-value for a one-tail test.

P-value = ________

Important things to remember about the *P*-value:

- When we assume that the null hypothesis is true, we're assuming that two *population* proportions are equal. Then we can think of the *P*-value as the chance that the observed difference in *sample* proportions is due to chance variation.
- When the *P*-value is small enough, we no longer believe that the difference between population proportions is likely to be zero. In this case, we reject the null hypothesis that the population proportions are equal. We say that our sample data supports the alternative hypothesis.
- When the *P*-value is too large, we do not reject the null hypothesis. We say that our sample data do not support the alternative hypothesis.
- We decide whether a *P*-value is "too large" or "small enough" to reject the null hypothesis by comparing it to our chosen significance level, α.

8 Interpret the *P*-value in the context of this problem.

Step 4: State a Conclusion

It is important to explain the result of your test in a way that is easy to understand. Remember that the hypothesis testing process is not perfect. So when you explain the result of your test, make clear that there is some uncertainty.

A conclusion can express support for the alternative hypothesis. The strongest support we ever give to the null hypothesis is to say that *we cannot disprove it*. It is always wise to indicate that the conclusion was based upon sample data.

9 For the exoneration rate test, make a decision using the *P*-value and α.

A What is the significance level?

α =__________

B Use the *P*-value to make a decision regarding the null hypothesis.

C Think about the exoneration rate data. State a conclusion in the context of the data.

NEXT STEPS

Here in the U.S., we have been hearing about the effects of global warming but we are interested in whether there has been any change in people's opinions about global warming. In two consecutive years, 2010 and 2011, the Pew Research Center conducted polls to find the proportion of adults in the U.S. who were concerned about the effects of global warming. The results of the two consecutive polls seemed to indicate that there had been a change in public opinion over this period of time.

Suppose that we decide to test the claim that the proportion of adults in the U.S. who were concerned about the effects of global warming changed between 2010 and 2011. We gather independent samples. In the initial year (2010), our data tells us that x_i = 233 out of n_i = 359 randomly sampled adults were concerned about global warming. The following year, x_f = 194 out of a random sample of n_f = 352 adults were concerned about global warming.

Using a 1% significance level, we can now test the claim that the true population proportions are different.

10 Step 1: Determine the Hypotheses

For the global warming test, we claimed that the proportions of people concerned about global warming were different for the two years under consideration.
Let p_i represent the proportion of adults in the U.S. who were concerned about global warming in the initial year. Then let p_f represent the proportion of adults in the U.S. who were concerned about global warming in the following year.

A What is the null hypothesis for this test?

B What is the alternative hypothesis?

C Is this a one-tailed or two-tailed test? Think about the null hypothesis. How do you know that this is a one or two-tailed test?

11 Step 2: Collect the Data

A Verify that the criteria for approximate normality are met.

For the global warming test:

B Calculate $\widehat{p}_i$ and $\widehat{p}_f$.

C Calculate the observed difference in sample proportions, $\widehat{p}_i - \widehat{p}_f$.

12 Step 3: Assess the Evidence

A Use technology to calculate $\overline{p}$, then the standard error, and finally the test statistic.

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} = \underline{\hspace{3cm}}$$

$$SE = \sqrt{\frac{\overline{p}(1-\overline{p})}{n_1} + \frac{\overline{p}(1-\overline{p})}{n_2}} = \underline{\hspace{3cm}}$$

$$Z = \frac{(\widehat{p}_i - \widehat{p}_f) - (p_i - p_f)}{s_{\widehat{p}_i - \widehat{p}_f}} = \frac{(\widehat{p}_i - \widehat{p}_f)}{s_{\widehat{p}_i - \widehat{p}_f}} = \underline{\hspace{3cm}}$$

B How many standard errors is the observed difference, $\widehat{p}_i - \widehat{p}_f$, from the *assumed* mean of differences of zero?

C Does the test statistic seem to support the claim that the proportion of adults in the U.S. who were concerned about the effects of global warming changed between 2010 and 2011? Why? How do you know?

D Use the test statistic to determine the *P*-value. Sketch the normal distribution, identify the position of the test statistic, and shade the area that corresponds to the *P*-value.

E Interpret the *P*-value in the context of this problem.

13 Step 4: State a Conclusion

A What is the significance level?

$\alpha =$__________

B Do you reject or do you fail to reject the null hypothesis?

C This hypothesis test focused on whether or not people's opinions on global warming have changed over the years. Write a conclusion to this hypothesis test. In your conclusion, explain what you believe the data suggests about opinions on global warming.

14 Use technology to compute a 90% confidence interval for the difference in population proportions, $p_i - p_f$.

A The critical value for the 90% level of confidence is Z_c = 1.645. Use technology to compute the margin of error for the estimate, $\hat{p}_i - \hat{p}_f$.

$$E = z_c \cdot \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i} + \frac{\hat{p}_f(1 - \hat{p}_f)}{n_f}} =$$

B What is the confidence interval for the difference in proportions, $p_i - p_f$?

$$(\hat{p}_i - \hat{p}_f - E,\ \hat{p}_i - \hat{p}_f + E)$$

15 Answer the following question based on this confidence interval:

Can we conclude which population was more concerned about global warming? That is, can we conclude that the people in 2010 were more concerned about global warming? Or can we conclude that the people in 2011 were more concerned about global warming?

Exercise 7.2-SJ

Part 1: Disciplinary Action in Schools

Are certain groups of students disciplined more severely than other groups of students? Researchers wanted to examine if student disciplinary procedures in a large school district show evidence of racial bias. The manner in which students are disciplined when they misbehave in school is a serious concern for advocates of social justice and equity. Excessive disciplinary action can have long-term negative impacts on students' performance and success in school and attitudes about school and education.

To investigate this issue, researchers collected disciplinary records of similar student misbehavior among students in the district. From the disciplinary records, they randomly selected 345 records of White male students and 318 disciplinary records of African-American male students. They observed that 36 of the White males and 55 of the African-American males were suspended or expelled.

We will test the claim that the proportion of White males who are suspended or expelled is lower than the proportion of African-American males who are suspended or expelled, even when the student misbehavior is comparable. We will test this claim at the 5% significance level.

1 **Determine the Hypotheses**

For this test, we have claimed that the proportion of White males, p_w, who are suspended or expelled *is lower than* the proportion of African-American males, p_a, who are suspended or expelled.

A What is the null hypothesis for this test?

B What is the alternative hypothesis for this test?

2 **Collect the Data**

A Determine the numbers of successes and failures in the two samples (White males and African-American males). Use these to check whether the approximate normality criteria have been met for both populations. In this context, a success is a student being expelled or suspended.

	Successes	**Failures**
White males	x_1 =	
African-American males	x_2 =	

B Have the approximate normality criteria been met for both populations?

C Compute the sample proportion for White males. Round to three decimal places.

$\hat{p}_w$ = __________

D Compute the sample proportion for African-American males. Round to three decimal places.

$\hat{p}_a$ = __________

E Compute the observed difference, $\hat{p}_w - \hat{p}_a$.

$\hat{p}_w - \hat{p}_a$ = __________

3 **Assess the Evidence**

A Compute the pooled proportion of all sample subjects who were suspended or expelled.

$$\bar{p} = \frac{x_w + x_a}{n_w + n_a} = \_\_\_\_\_\_\_\_\_\_$$

B Use technology to calculate the standard error. Round to three decimal places.

$$SE = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_w} + \frac{\bar{p}(1-\bar{p})}{n_a}} = \_\_\_\_\_\_\_\_\_\_$$

C Use technology to calculate the test statistic. Round to three decimal places.

$$Z = \frac{(\hat{p}_w - \hat{p}_a) - (p_w - p_a)}{SE} = \_\_\_\_\_\_\_\_\_\_$$

D Do you consider the observed difference to be unusually far from the assumed mean of differences?

E Use the test statistic to determine the *P*-value.

4 **State a Conclusion**

A What is the significance level for this test?

$\alpha =$__________

B What conclusion should we make about the hypotheses?

C Using complete sentences, write a conclusion explaining what you have learned about whether the sample data supports the claim.

Part 2: Differing Views on Interactions with Police

A recent Gallup poll[1] found differences in the ways in which white and Hispanic Americans viewed their recent interactions with police. Data on citizen interactions with police are collected and analyzed to understand how groups of Americans experience policing. These studies can provide useful information to support the need for policing reform. The simulated data in this problem are based on the results of the Gallup poll.

Suppose we wish to examine if there *is a difference* between the proportions of white and Hispanic Americans who believe that they were treated fairly by police during their most recent interaction. From a random sample of 66 whites 61 say they were treated fairly by police. From a random sample of 58 Hispanics, 49 say they were treated fairly by police.

Test the claim that the proportion of white Americans, p_w, who feel they were treated fairly by police *is different* than the proportion of Hispanic Americans, p_h, who feel they were treated fairly by police. Use a 1% significance level.

1 **Determine the Hypotheses**

A What is the null hypothesis for this test?

B What is the alternative hypothesis for this test?

2 **Collect the Data**

A Use the sample proportions to verify the criteria for normality for each of the underlying sampling distributions.

For the white Americans, there are __________ successes and __________ failures.

For the Hispanic Americans, there are __________ successes and __________ failures.

B Are the criteria for approximate normality met?

[1] https://news.gallup.com/poll/316247/black-americans-police-encounters-not-positive.aspx

3 **Assess the Evidence**

A Compute the pooled proportion. Round to three decimal places.

$$\bar{p} = \frac{x_w + x_h}{n_w + n_h} = \_\_\_\_\_\_\_\_\_\_\_\_$$

B Calculate the standard error. Round to three decimal places.

$$SE = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_w} + \frac{\bar{p}(1-\bar{p})}{n_h}} = \_\_\_\_\_\_\_\_\_\_$$

C Calculate the test statistic. Round to three decimal places.

$$Z = \frac{(\hat{p}_w - \hat{p}_h) - (p_w - p_h)}{SE} = \_\_\_\_\_\_\_\_\_\_$$

D How many standard errors is the *observed* difference, $\hat{p}_w - \hat{p}_h$, from the assumed *mean of differences* of zero?

E Use the test statistic to determine the *P*-value. Remember, you need to compute *P*-values for two-tailed hypothesis tests differently.

4 **State a Conclusion**

A What is the significance level for this test?

B Consider the *P*-value and the significance level. Make a decision regarding the null hypothesis.

C Using complete sentences, write a conclusion in context.

5 Use technology to compute a 99% confidence interval for the difference in population proportions,

$p_w - p_h$.

A A 99% level of confidence allows Z_c = 2.576 standard errors in the margin of error. Calculate the margin of error in the estimate, $\hat{p}_w - \hat{p}_h$.

$$E = Z_c \cdot \sqrt{\frac{\hat{p}_w(1 - \hat{p}_w)}{n_w} + \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h}} =$$

B Construct the lower bound confidence interval for the difference in population proportions, $p_w - p_h$.

$$\hat{p}_w - \hat{p}_h - E$$

C Construct the upper bound confidence interval for the difference in population proportions, $p_w - p_h$.

$$\hat{p}_w - \hat{p}_h + E$$

D Does your confidence interval permit the proportions of whites and Hispanics who feel they were treated fairly by police to be equal? Explain.

E Does your confidence interval agree with the conclusion of the hypothesis test above? Explain.